# IS-ENES3 Deliverable D7.1

## First KPI and TA report for ENES CDI data services

*Reporting period: 01/01/2019 – 30/06/2020*

*Dissemination level: public*

Authors:
Stephan Kindermann, Maria Moreno, Martina Stockhause (DKRZ)
Martin Juckes (UKRI)
Guillaume Levavasseur (CNRS-IPSL)
Sandro Fiore, Alessandra Nuzzo (CMCC)
Alessandro Spinuso (KNMI)
Lars Bärring (SMHI)

Reviewer(s): Sylvie Joussaume and Sandro Fiore
Release date: 30/07/2020

**ABSTRACT**

The ENES CDI services encompass data search, access, and support services (Task1), associated processing services (Task2 - VA and Task3 - TA), data standards services (Task4), and support for CMIP6 documentation (Task5). In this deliverable we summarize characteristic usage information for each service category and provide associated statistics. The service provisioning is based on eleven installations distributed across Europe. The installation specific details are provided as part of the associated IS-ENES3 Reporting Period 1 (RP1) report. The evolution of the services is coordinated in cooperation with WP5/NA4 and WP10/JRA3. Sustainability aspects of the services are contributed and discussed further in the sustainability work package WP2/NA1.

| Revision table | | | |
|---|---|---|---|
| **Version** | **Date** | **Name** | **Comments** |
| Release for review | 22/07/2020 | Stephan Kindermann, Maria Moreno (DKRZ) | |
| Final | 30/07/2020 | Stephan Kindermann, Maria Moreno (DKRZ) | Reviewers comments integrated |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

# Table of contents

## Executive Summary

The ENES CDI data and metadata services are based on eleven installations across Europe closely collaborating to provide a consistent set of data distribution, access, and associated support services on the one hand and critical support for data standardization and model documentation on the other hand. Larger ENES CDI sites additionally collaborate to provide a set of compute services (VA and TA) which are associated with their large climate model data pools thus providing a distributed European climate model data base supporting model data evaluation and processing activities.

For each of the services provided by the ENES-CDI a set of key performance indicators was defined. A summary and overview is given in Table 1 in Section 2. Subsection 2.1 summarizes the data access statistics for the ESGF data distribution and user support services. It also includes associated service statistics for the support of persistent data identification and data citation. Subsection 2.2 characterizes the new emerging VA compute services, which will be included in future reporting periods. The newly established transnational access (TA) service providing virtual workspaces at the installations at DKRZ, CNRS-IPSL, UKRI and CMCC is described in Subsection 2.3, together with the formal selection and review procedure. The support activity for data standards (CF and data request) is described in the Subsection 2.4, whereas the support activity for model documentation (es-doc) is summarized in the Subsection 2.5.

A new ESGF Data Statistics service (formerly ESGF Dashboard) with a User Interface was established showing ENES-CDI data distribution KPIs[1].

## 1. ENES CDI data and metadata services: Objectives and Overview

Users from the climate modeling and impact research communities as well as users from interdisciplinary domains rely on stable and consistent services to access and process high volume climate model data from the World Climate Research Programme (WCRP[2]) reference simulations. The core of the WCRP reference simulations are the Coupled Model Intercomparison Project (CMIP) and the Coordinated Regional Climate Downscaling Experiment (CORDEX) model data, which are in distributed repositories. The international repositories partnership is the Earth System Grid Federation[3] (ESGF), with IS-ENES supporting the European contribution.

[1] http://esgf-ui.cmcc.it/esgf-dashboard-ui/isenes3-kpi.html
[2] https://www.wcrp-climate.org/wgcm-overview
[3] https://esgf.llnl.gov/

The ENES Climate Data Infrastructure (ENES CDI) provides: (1) access services on CMIP and CORDEX data from the Earth System Grid Federation (ESGF), the archival system (the World Data Center for Clima, WDCC), and the Climate4Impact portal, (2) processing services, and (3) services on data documentation and standards. These services are mainly offered through virtual access (VA). Users have also the possibility to apply for virtual workspaces through a trans-national access (TA), which allows them to remotely access not only the data pools but also the IS-ENES3 computing infrastructure (high performance computers (HPC) and clusters) hosting the data.

The goal of these service activities is to provide operational support to the climate and climate impact research communities and other communities using the data and tooling provided by IS-ENES3. The VA/TA activities will also provide support for communities that are new to using climate data.

## 2. Service statistics and performance indicators (PIs)

PIs and Key PIs (KPI) summary table:

| | |
|---|---|
| ESGF data download KPIs and PI | KPI : Number of downloads (EU/no-EU/no geo-located) |
| | KPI: Downloaded data volume (EU/no-EU/no geo-located) |
| | KPI: Number of distinct users (EU/no-EU/no geo-located) |
| | KPI: Number and percentage of emails answered in the user support mailing list by an ENES member |
| | PIs: CORDEX specific number of downloads, volume, and distinct users and number of answers to the new CORDEX user support mailing list |
| Replication and archival PIs | Number of TB of original data |
| | Number of TB of replicated data |
| | Number of TB of overall volume |
| Data citation PIs | Number of registered DOI registered to DataCite |
| | Number of revisions of citation information published to DataCite |
| | Number of citation entries added to the service database |
| Persistent | Number of original and number of replica CMIP6 datasets |

| | |
|---|---|
| identifiers PIs | Number of original and number of replica CMIP6 files |
| WDCC PIs | Number of downloads |
| | Downloaded data volume |
| | Number of distinct users (EU/no-EU/no geo-located) |
| Climate4Impact KPIs and PIs | KPI: Unique Users |
| | KPI: Number of access to the users' personal space (Basket Requests) |
| | PI: Number of map visualisations requested by users (WMS Get Map Requests) |
| | PI: Number of processing functions executed by users (WPS Execute Requests) |
| | PI: Number of data subsetting requests by users (WCS GetCoverage Requests) |
| | PI: Number of hits |
| CF data model PI | Release of package updates |
| CF Standard Name PIs | Publication of new versions of the table |
| | New terms published |
| CMIP Data Request PIs | Issues resolved |
| | Releases |
| ES-DOC PIs | Issues registered on the web service |
| | Number of documentation search and web site visits |
| | Number of questions to the helpdesk |
| | Metadata generated by the cdf2cim process of the ESGF publisher |

Table 1. KPIs and PIs summary table.

## 2.1 ESGF data dissemination, data archival and Climate4Impact services and user support (Task1)

The ESGF based data dissemination is coordinated across data node installations at partner sites and data portal installations at large replica sites (DKRZ, UKRI and CNRS-IPSL) which are all tightly integrated into the worldwide ESGF data federation. The European contribution to the overall CMIP6 ESGF effort is provided based on statistics on number of downloads, downloaded data volume as well as number and geographical distribution of users together with the associated user support activities (see Subsection 2.1.1). Replication and archival statistics are shown in Subsection 2.1.2. The ESGF citation service for CMIP6 is completely provided as part of the ENES CDI. The number of DOI citation references and the number of support requests is in Subsection 2.1.3.

A dedicated impact community targeted portal, the Climate4Impact portal[4], is hosted at KNMI and a new version is currently in active development. Some related user activity indicators are shown in Subsection 2.1.4.

### 2.1.1 ESGF data download KPIs and statistics and user support statistics

Four KPIs have been developed regarding the data download and user support metrics to give a better understanding of the amount of data exploited via IS-ENES3/ESGF by the data users all over the world. We report the same ESGF KPIs as in the last IS-ENES2 report[5] for the RP1. They comprise the information of aggregated usage statistics from data provided by several projects, being CMIP6, CMIP5, and CORDEX, followed by input4MIPs, the most popular ones.

**ENES CDI ESGF data download KPIs**

The three KPIs in Fig. 2.1 quantify the monthly number of files downloaded from the EU ESGF data nodes and the associated data volume (with a distinction between complete and partial downloads, both considered as successful downloads), as well as the monthly number of distinct users successfully performing the downloads. The first two KPIs are updated on a weekly basis via the ESGF Data Statistics service[6] user interface (developed at CMCC within WP10/JRA3) which shows data usage metrics. The distinct users KPI, instead, cannot be automatically handled by the statistics tool due to sensitive information and it is updated every six months. The term "*users*" is used for consistency with the KPIs defined in the previous phases of IS-ENES;

[4] https://climate4impact.eu/
[5] https://is.enes.org/archive-1/phase-2/documents/contractual-documents/kpis-report/view
[6] http://esgf-ui.cmcc.it/esgf-dashboard-ui/

however, due to the EU General Data Protection Regulation (GDPR) and the new CMIP6 open data policy, by monthly distinct users we mean the "*average number of monthly distinct clients per data node*". With respect to the previous two KPIs (number of files and data volume), the distinct users metric is non-additive, which explains why we calculated the average instead of the total. Additionally, it should be noted that it provides a lower bound for the actual number of distinct users. These KPIs are measured for clients in Europe, outside Europe, and non-geolocated clients (sometimes it is not possible to convert an IP address used for the request, because that IP could belong to a local network or it is not registered into one the different databases used by the geo-location services; instead of excluding such IPs, we provide them as well, but distinguishing them from the other classification). Besides these KPIs, the new User Interface of the ESGF Data Statistics service provides a comprehensive set of data download and data publication metrics.

| | EU | Not EU | Non geo-located | Total | Monthly average |
|---|---|---|---|---|---|
| Number of downloads (files) | 10748428 | 68620942 | 749797 | 80119167 | 4451065 |
| Downloaded data volume | 2.97 PB | 4.87 PB | 159 TB | 8 PB | 455 TB |
| Number of distinct users[7] | 269 | 831 | 21 | - | 1121 |

Table 2: Data download KPIs & Distribution

Overall numbers for the 3 IS-ENES3 data download KPIs during the RP1 (see Table 2) are:

1. A total number of 80119167 files has been downloaded from the European IS-ENES3 data nodes during the period from January 2019 to June 2020. These downloads are divided as follows: 10748428 downloads coming from users in the EU, 68620942 downloads from users outside the EU and 749797 downloads from non-geolocated users. On an average of 18 months, a number of 4451065 downloads occurred each month, divided as follows: 597135 downloads coming from users in the EU, 3812275 downloads from users outside the EU, and 41655 downloads from non-geolocated users.
2. Correspondingly, a total data volume of 8392407 GB (8 PB) has been downloaded, distributed as follows: 3116622 GB (2.97 PB) from users in the EU, 5112608 GB (4,87 PB) from users outside the EU and 163199 GB (159 TB) from non geolocated users. Over

[7] Mean number of monthly distinct clients per data node averaged over the reporting period of 18 months.

the reporting period of 18 months, an average number of 466245 GB (455 TB) has been downloaded, divided as follows: 173146 GB (169 TB) from users in the EU, 284034 GB (277 TB) from users outside the EU, and 9067 GB (8,85 TB) from non-geolocated users.

3. As concerns the distinct users, an average of 1121 distinct users performed downloads from the European data nodes. This figure is the result of the mean number of monthly distinct users per data node averaged over the reporting period of 18 months. They are divided as follows: a monthly average number of 269 for users in the EU, 831 for users outside the EU and 21 for non-geolocated users.

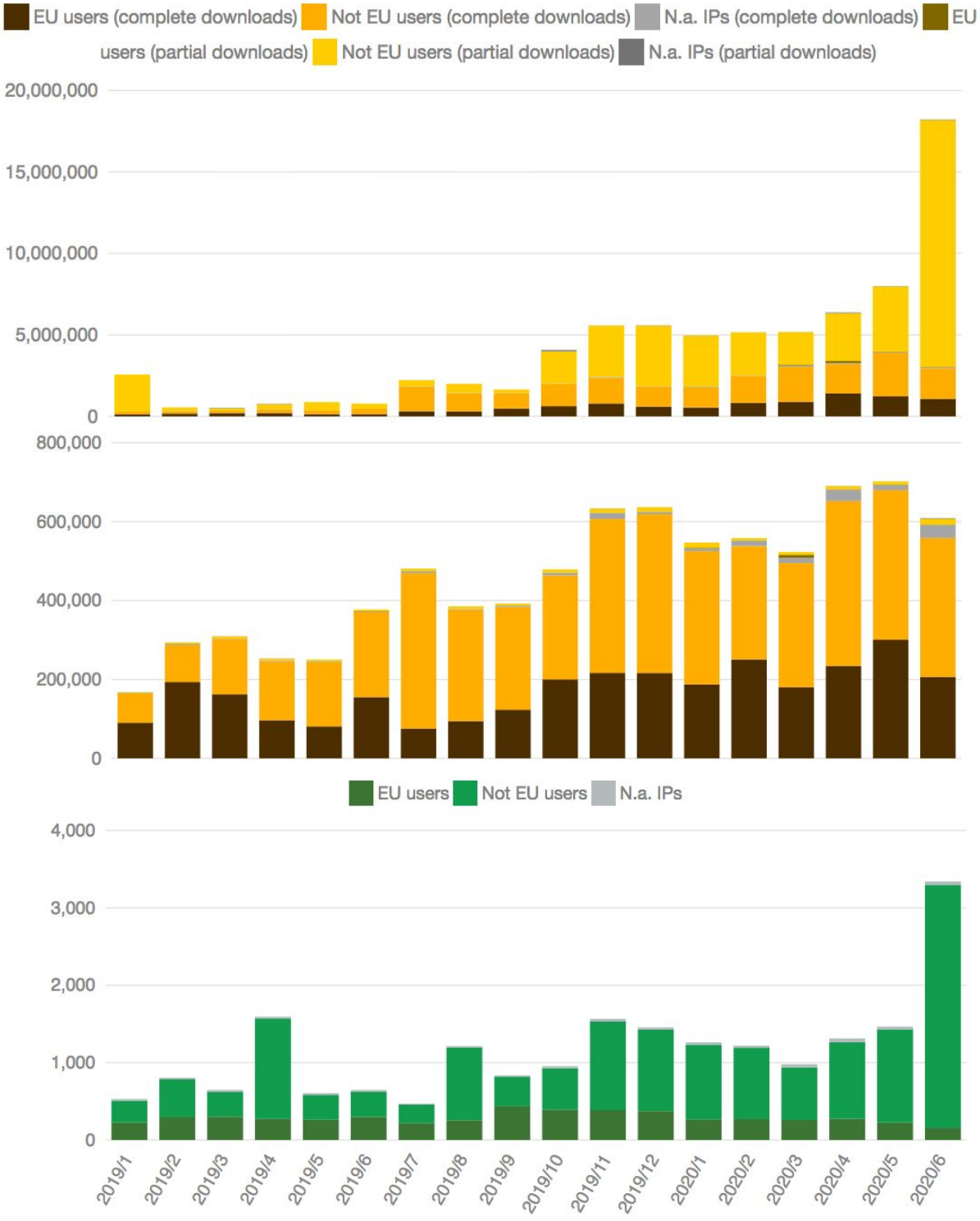



Figure 2.1. ESGF data download KPIs (stacked charts): number of downloaded files (top), associated data volumes (in GB) (center), and distinct users per node (bottom).

**ESGF user support**

The fourth KPI quantifies the IS-ENES3 contribution to the users support mailing list (ESGF help desk). The ESGF user enquiries on the ESGF user mailing list are managed by the IS-ENES3 support team in collaboration with other non-EU ESGF participant institutions (see Fig. 2.2). During the RP1, the ESGF user support mailing list received a total of 1485 emails mainly from CMIP and CORDEX users. Taking away the 277 unanswered emails (due to the lack of available supporters, some unfiltered announcements, or questions non related to data management issues but model performance, variables, scientific meaning, etc. which are forwarded to a second level support[8]), there were 526 threads (which corresponds to about 2 distinct requests per working day), from which 365 (~ 70%) were solved with the help of an IS-ENES3 supporter. The Fig. 2.2 shows the statistics per month and in Annex A is the evolution between project phases. Note that "All" stands for answered and unanswered emails and the percentage is over the total number of threads, that is, including the unanswered emails.

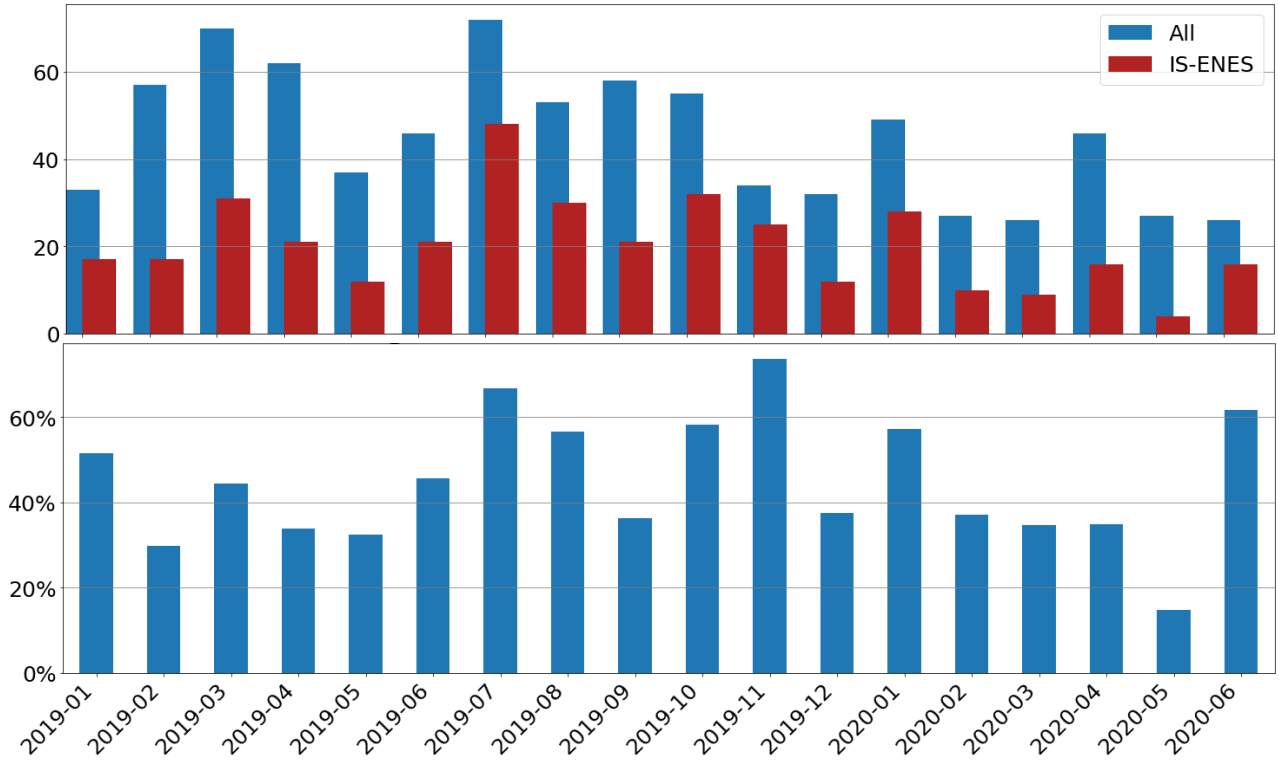



Figure 2.2. ESGF user support activity: number of support email threads (top) and the percentage of threads with the contribution of an IS-ENES3 supporter (bottom).

[8] WIP mailing list wgcm-wip@wcrp-climate.org or the CORDEX specific support, see next page.

The European ENES CDI partner sites cooperate and coordinate the European support activity based on a first/second level support structure. Given the increase of model data volume and complexity with CMIP6, user support requires a significant effort, also because there is no dedicated overall ESGF coordinated technical support system(s) (support forum, ticketing system, chat bot, etc.). The new open access policy for CMIP6 reduced the data user access problems significantly in comparison to CMIP5.

Beside the user support mailing list (helpdesk), the WP7 is a significant contributor to the other activities of the ESGF User Support and Documentation Team. In particular, DKRZ maintains the ESGF user tutorial and FAQS[9] (see M7.1 and Annex B for a list of updates), and reviews and coordinates the unification and update of the ESGF portals look & feel (see M7.1 and the Annex C for a list of updates). DKRZ and CMCC, in collaboration with WP3/NA2, created training material and disseminated ESGF data service activities.

Specific CORDEX support is being developed to cover the scientific needs of the corresponding users community since the ESGF general user support mainly helps in data management related issues (where to find, how to download,...) but the help desk barely answers questions on models performance, variables, scientific meanings, etc. The CORDEX specific support has focussed on three main actions:

1. To respond to user requests, mainly via email for support regarding data, and based on the experiences from this to better organise the support activities for improved efficiency and user experience. To unify the multiple sources of user request, a new CORDEX user support mailing list[10] and a new FAQ[11] have been established and disseminated.
2. To enable the main CORDEX website[12] to show near real-time statistics of all CORDEX data served through the ESGF Data Statistics service.
3. To support the regional modelling community in preparation of the CORDEX 2nd Phase by organising the data request specifying the standards for model output. In preparation of CORDEX Phase 2, work is ongoing for defining the exact technical specification of regional model output. This builds on, and is a simplification of the corresponding work on a Data Request for CMIP6. In essence this activity means engaging with the regional modelling institutes to facilitate the interaction leading to a commonly agreed flexible document specifying the output standard**.**

[9] https://esgf.github.io/esgf-user-support/index.html#
[10] datasupport@cordex.org
[11] https://cordex.org/faq/
[12] www.cordex.org

**CORDEX PIs**

Since the early days of CORDEX there have been repeated requests for near real-time statistics of data available via ESGF, both the volume of data available and amount of data downloaded, divided/aggregated per region, global forcing model, regional model, popular variables etc. Until now such statistics have been difficult to acquire and required substantial efforts by ESGF node managers. We have been in extensive contact with the developers of the new User Interface (UI) of the ESGF Data Statistics service where CORDEX needs have been used as a use case. In this respect, a CORDEX view has been integrated into the ESGF Data Statistics service UI with a set of statistics regarding data download in terms of files and data volume[13]. Work is underway to integrate this kind of statistical displays into the main CORDEX website. This will support potential users and stakeholders in getting an overview of data holdings, international funding organisations and the WCRP to get an overview of the usage of the output of the whole CORDEX activity. PIs will be the CORDEX specific number of downloads, volume, and distinct users, and number of issues solved in the new CORDEX user support mailing list, and they will be reported in the next reporting periods.

### 2.1.2 Replication and archival statistics

The ENES CDI partners DKRZ, UKRI and CNRS-IPSL maintain coordinated large model data pools to provide resilient and efficient access to CMIP data collections to the European research community. These pools contain original data from European modeling centers as well as replicas collected from remote ESGF sites across the world. These pools build the basis for the ESGF data distribution services and provide the core data basis for the provided VA and TA compute services. In case the users request data from an ESGF node that is temporarily down (usually due to maintenance), replication allows them to reach the same data in one of the larger ENES CDI sites from the same portal. This valuable service is used on a daily basis although there is no metric to provide qualitative information on the usage (ESGF data download KPIs in Fig. 2.1 do not distinguish between original and replicated data provision). Table 3 shows the amount of data replicated at ENES-CDI ESGF Tier-1 sites an overview of the currently available CMIP6 data volumes collected and provided.

| ENES data center | Original data (PB) | Replicated data (PB) | Overall volume (PB) |
|---|---|---|---|
| DKRZ | 1.5 | 2.6 | 4.1 |
| CNRS-IPSL | 2.5 | 1.5 | 4 |
| UKRI | 1.1 | 0.7 | 1.8 |

Table 3. Volume of CMIP data pools at ENES-CDI ESGF Tier-1 sites.

[13] http://esgf-ui.cmcc.it/esgf-dashboard-ui/cordex.html

### 2.1.3 Data Citation and Persistent identifier service

The Citation service provides DOI registrations and updates of citation information, curation of citation information and user support for the projects CMIP6 and input4MIPs.

- DOI registrations and citation information updates (see Fig. 2.3): in the first half of 2020, 572 DOIs have been registered and 1 290 revisions of citation information have been published to DataCite. CMIP6 data providers started to significantly release data in July 2019.
- Metadata Curation: the metadata is based on the CMIP6 participant registration information (managed in the CMIP6 controlled vocabulary[14]), 854 citation entries have been added to the database of the service.

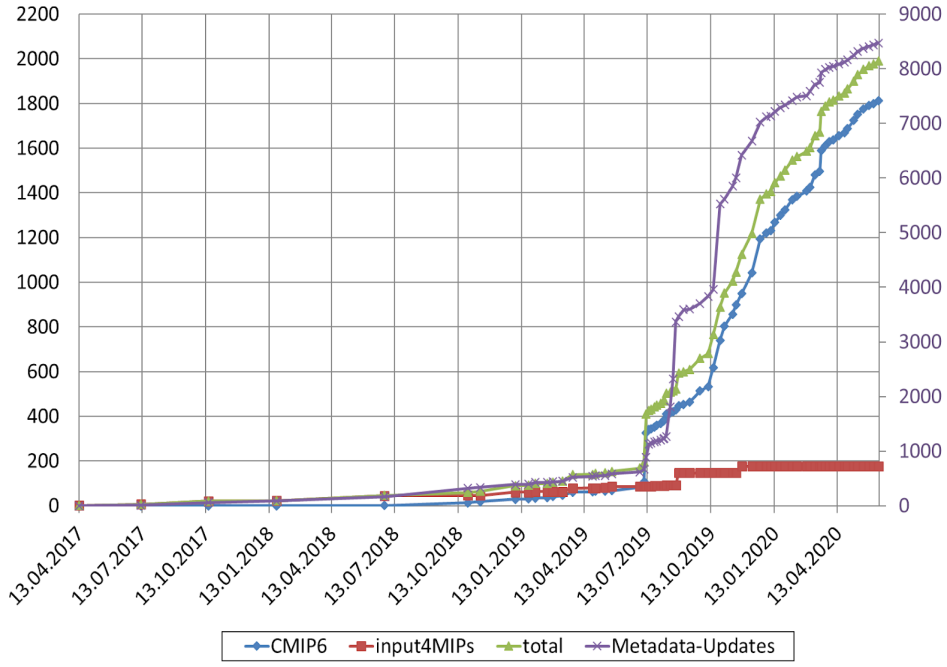



Figure 2.3 Number of DataCite DOI data citation registrations (left y-axis) and associated metadata updates (right y-axis) for CMIP6.

**PID service for CMIP6**

The Persistent Identifier (PID) service, which was integrated into ESGF in support of CMIP6, provides persistent references to files and datasets and supports stable references for versioning and replicas. Table 4 shows the PID registered for the CMIP6 project since the data providers released the data (about July 2019). There was not significant activity in the first semester of the RP1. In subsequent reporting periods, we will compare the cumulative statistics.

[14] https://github.com/WCRP-CMIP/CMIP6_CVs/

| | Total number | Original | Replica |
|---|---|---|---|
| Datasets | 6316650 | 3254555 | 3062095 |
| Files | 23524472 | 14373611 | 9150861 |

Table 4. Persistent Identifier Assignments in support of CMIP6.

### 2.1.4 WDCC statistics

The World Data Center for Climate (WDCC) provides data archival services in support of CMIP for the ESGF users, IPCC Data Distribution Center[15] (DDC) users and other WDCC associated user groups. Data is replicated to the hosting institution (DKRZ) and then ingested in the long term archive along with its describing metadata. The archived data is then re-integrated into the ESGF on a dedicated data node (included in the statistics in Fig. 2.1) serving data from tape.

During the reporting period, more than 1.9 million of files were downloaded from the CMIP5 long term archival node at DKRZ, which corresponds to 260 TB of model data, and 2.6 PB of CMIP6 data were replicated adding to the 1.5 PB original CMIP6 data hosted at DKRZ (see Table 3) and will be transferred in the long term archive in the next reporting periods. Figure 2.4 shows the PIs from the WDCC Climate and Environmental Data Retrieval and Archive (CERA) Interface[16].

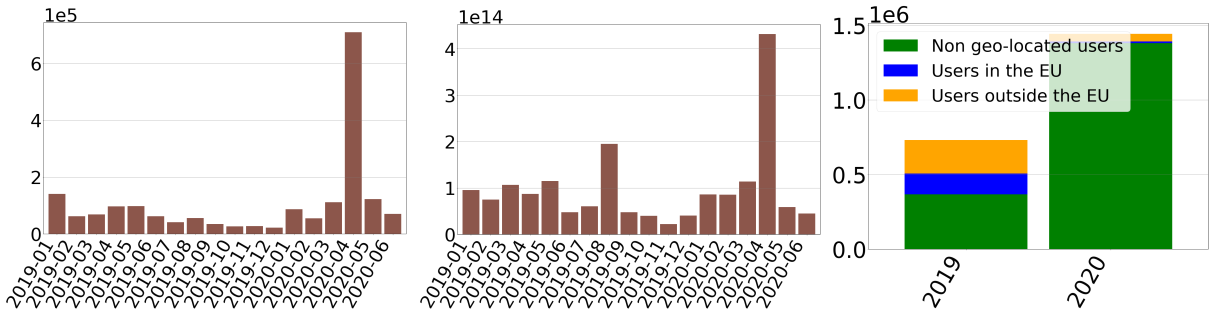



Fig 2.4 WDCC CERA PIs: number of downloaded files (left), associated volume in bytes (center), and users (right).

[15] http://www.ipcc-data.org/
[16] https://cera-www.dkrz.de/WDCC/ui/cerasearch/statistics_index

### 2.1.5 The Climate4Impact portal: access statistics and user support

**Climate4Impact access statistics**

For the Climate4Impact portal, the last KPIs report of IS-ENES2[17] consisted of the number of visitors, unique visitors, visited pages, hits, and downloaded bytes. The Climate4Impact portal is one of IS-ENES3 newer services and undergoes current upgrades, and consequently the KPIs need calibration. Several metrics have been improved and extended to provide the following two new KPIs that will be monitored and reported in the future reporting periods (see Table 5):

- Unique Users, that is, the actual number of registered users who accessed the service in the month (this is different from the Unique Visitors in previous reports, which instead reported any type of daily unique access, summed up for the whole month).
- Number of access to the users' personal space (Basket Requests).

Besides, the following are representative performance indicators (see Table 5):

- Number of processing functions by users (WPS Execute Requests).
- Number of map visualisations requested by users (WMS Get Map Requests).
- Number of data subsetting requests were done by users (WCS GetCoverage Requests).
- Number of access to the users' personal space (Basket Requests).
- Number of hits.

| | PIs | | | | KPIs | |
|---|---|---|---|---|---|---|
| | Hits | wps Execute Requests | wms Get Map Requests | wcs Get Coverage Requests | Basket Requests | Unique Users |
| Jan 2020 | 257.394 | 563 | 4541 | 38 | 5114 | 36 |
| Feb 2020 | 213.936 | 439 | 3255 | 1 | 4316 | 36 |
| Mar 2020 | 251.330 | 995 | 5435 | 0 | 5182 | 41 |
| Apr 2020 | 232.370 | 750 | 10549 | 7 | 4195 | 33 |
| May 2020 | 460.159 | 950 | 5820 | 105 | 6366 | 46 |
| June 2020 | 89.727 | 253 | 1850 | 36 | 2315 | 44 |

Table 5. Several indicators of the user activity at the Climate4Impact portal in 2020.

[17] https://is.enes.org/archive-1/phase-2/documents/contractual-documents/kpis-report/view

We specifically divided the statistics by the different services hosted by the portal. The standard OGC interfaces address different purposes and modes of access, in particular we use WPS, WMS, WCS[18] for data processing, visualisation, and coverage. These are directly exploited by the Climate4Impact portal on behalf of the users. We believe that such statistics are proving better insight on the functionalities that are more popular and on the actual impact and exploitation of the standard, suggesting targeted further improvements. We do not report page views as in previous states, but the actual use of the services. The total quantity of data downloaded through the portal is missing in the current figures. However we aim at having our users to download less data and to process more on remote resources (i.e. the WPS Execute Requests). In the future, we aim to thereby improve the monitoring of the amount of data processed remotely rather than how much users get into their laptop. The following statistics has not been monitored yet:

- Users geo-location: we will probably omit this or have to make estimates based on IP addresses, a functionality that has not been implemented yet.
- Data amount downloaded/used in processing per user: as data is not downloaded *through* Climate4Impact (we refer to the URL) this metric is hard to resolve and will be done in the current metrics, as it is not easily derived from the log files. We need to add specific log actions to the search and processing services (under development).
- Involved variables, models, projects, datasets, experiments etc.

In October 2018 the Climate4Impact machines had a severe malfunction and the service had to be moved to the Cloud (AWS). At that point in time Climate4Impact was not easy to deploy, it took approximately 1.5 month to get the service up and running again. The current Climate4Impact is optimized for easier deployment and has been redeployed on different locations (Surf Sara Cloud, AWS) before being deployed on its destination where it is running now. Deployment now takes days (including testing) instead of 1.5 months. Due to moving around, the 2019 log files needed to derive statistics are not available.

From early 2020 the Climate4Impact operational environment is deployed at AWS. Since this period we keep statistics, which provide more insight than before (AWStats, see Table 4). The current statistics show the user actions done (so actual usage) instead of visits. The number of hits is the only KPI that was also monitored during the IS-ENES2 phase and beyond. In 2018 (the year between IS-ENES2 and 3 phases) we averaged 179.871 hits per month and in 2020 the average per month is 249.504 up to now, which represents a 38% increase.

[18] https://www.ogc.org/standards/wps , https://www.ogc.org/standards/wms , and https://www.ogc.org/standards/wcs

**Climate4Impact user support**

For user support Climate4Impact closely cooperates with actions in WP3/NA2. User support was provided by co-organizing two user workshops at the EGU2019[19] together with WP3/NA2. In these workshops we targeted impact users and explained to them how they can use climate model data in their research and how C4I can help them to select the data they need. We will also support the IS-ENES3 Autumn Summer school (Prague) organized by WP3/NA2 by providing the new search, download and indicator processing tooling and support during the summer school. Until now, a very low number of users have sent questions via the contact form[20].

### 2.2 Compute service: derived data products and web services (VA, Task2)

The WP7 established a new service ("Comp1") offering easy direct access to different compute facilities and/or pre-calculated derived data products at dedicated ENES CDI partner sites. This service will also provide a lightweight access procedure (via VA) for user communities to test service offerings of the ENES CDI with respect to data associated compute possibilities and thus also serves as a preparatory platform to plan later TA service applications. Since a wide variety of data users is currently targeted and a variety of services will be provided which are partially in a pre-production and test status, no detailed usage statistics are available so far:

- Derived data product generation and provisioning:
  a pre-operational service was deployed to make pre-calculated climate evaluation results available to the broader climate science community. A first test service is deployed at DKRZ and uses the ESMValTool to calculate the results based on the large CMIP6 replica data pool at DKRZ. A wide range of users around the world already used this service and thus this service will now be stepwise operationalized and more detailed usage statistics will be available in the next periodic reports.

- Web processing service provisioning including dedicated compute services for integration with the Climate4Impact portal: initial test deployments are available at DKRZ as well as UKRI and CNRS-IPSL.

This VA service will also provide important information to plan and focus future compute service offerings. As already noted above, currently parallel routes are explored:

[19] https://meetingorganizer.copernicus.org/EGU2019/session/33654 and https://meetingorganizer.copernicus.org/EGU2019/EGU2019-11077.pdf
[20] https://climate4impact.eu/impactportal/help/contactexpert.jsp

1. the provision of web accessible access points for processing near the data, which includes interactive access points (e.g. via jupyter notebooks) as well as programmatic APIs (based on the OGC web processing service standard),
2. the provision of interfaces to pre-calculated derived data products, which were generated based on the ESMValTool. In this context a first service prototype has been deployed and made accessible to a broader user community, and
3. the provision of dedicated processing services for the climate for the impact community.

### 2.3 Virtual workspaces (TA, Task3)

**Short description of the components of the TA service**

New data-near server-side processing capabilities at the IS-ENES3 HPC centers are now made accessible to a broader user community via the new service providing access to Analysis Platforms using the TA funds ("Comp2" services).

Data-near server-side processing workflows significantly reduce the typical issues occurring in classical client-side workflows: heavy data transfer, strong memory requirements, and slow sequential processing. Thanks to the Analysis Platforms, users can run their own scripts in their own environments on the IS-ENES3 High Performance Computing (HPC) centers, which are associated with large efficiently accessible climate model data pools. Thus users do not have to download the data and can speed their analysis by using parallel computing. This new server-side data-near service is therefore especially beneficial for data users willing to compare multiple model data, as well as those with a low network bandwidth or limited access to high performance facilities.

**TA providers**

Four IS-ENES3 HPC centers are providing access to Analysis Platforms. All the providers offer:

- Computational and storage resources,
- Tools, libraries and services of the classical set of software used for climate model analysis. Additional environments and virtual machines can also be launched on demand,
- A set of collections of climate model data for climate model intercomparison experiments. Data can be replicated and kept in sync with the ESGF federated data archive on users demand, all the providers have allocated space for this purpose. The data pool is efficiently accessible from the shell as well as from Jupyter Notebooks. All the variable-centric collections are browsable at any ESGF portal.
- Documentation and user support.

Each installation offers the following specific features during the whole IS-ENES3 project:

- DKRZ-Comp2
    1. 3.2 petabytes of CMIP6 data and 1.2 PBytes of CMIP5 and CORDEX, besides some CMIP related projects (Obs4Mips,..), all data browsable at the specific catalogue[21].
    2. 43 pre and post processing nodes with 2x12 cores each (Intel Xeon E5-2680 v3 (Haswell) @ 2.5GHz), and 32000 CPU hours on the HPC system (with 3.300 compute nodes encompassing more than 100.000 compute cores accessible via a Slurm batch system), and
    3. 100 TByte of storage.

- CNRS-Comp2 (at CNRS-IPSL)
    1. 1,5PByte of CMIP and CORDEX data and observational datasets (Reanalysis, Obs4MIPs, input4MIPS,...),
    2. a cluster with compute nodes (2000 cores, up to 256 GB RAM/node, PBS resource manager, Network (InfiniBand QDR à 40Gbits/s and FDR at 56Gbit/s), and 31000 CPU hours, and
    3. 50 TB shared storage for data analysis (Lustre), temporary and final results.

- CMCC-Comp2
    1. 96 TB CMIP5 and 50 TB CMIP6,
    2. analytics cluster with 5 compute nodes (20 cores/node, 256 GB RAM/node, 1TB local disk each), SLURM resource manager, Network 10 Gbit/sec (10 GbE), 30TB storage (GFS) and a HPC cluster based on 348 Lenovo SD530 biprocessor nodes (for a total of 12.528 cores with a theoretical peak performance of 1.2 PetaFlops), Infiniband EDR network, and LSF scheduler, and 34000 CPU hours, and
    3. 50 TB shared storage for data analysis (GPFS).

- UKRI-Comp2 (TA at JASMIN-CEDA)
    1. access to 10PB of archived data, with thousands of data collections, including CMIP6 simulation results, reanalysis datasets, and large satellite datasets, browsable in an specific catalogue[22],
    2. up to 12,000 compute cores, and 24000 CPU hours.
    3. 50TB of storage for data analysis.

[21] https://cmip-esmvaltool.dkrz.de/?login_required=1&next=/solr/data-browser/
[22] http://archive.ceda.ac.uk/

**TA service management**

We defined the timeline for the application and selection procedure, two calls per year was the most suitable frame. We homogenized the service among different access providers since some of them already provided similar services to external users and others do not.

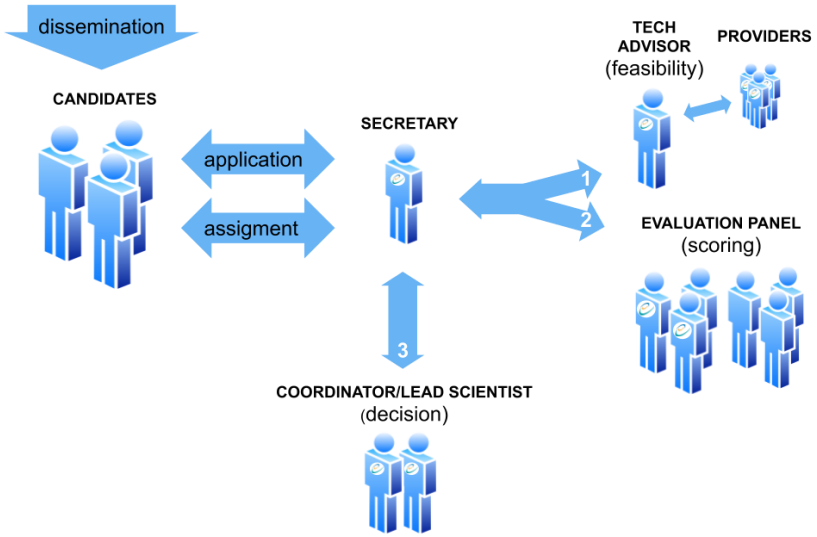



Figure 2.5 Description of the selection workflow.

We designed the selection procedure following the guidelines in the H2020 GA Article 16. We decided that the best is a four steps review: administrative, technical, scientific and a final assignment. After the administrative revision, the selection is divided in 3 steps (see Fig. 2.5):

1. TECH ADVISOR: A providers' representative (rotational). First and second call: Dr. Stephan Kindermann (DKRZ). All the providers evaluate the technical feasibility of the appropriate use of resources. Proposals should demonstrate awareness of the advantages of the infrastructure that they wish to use; they must explain how their project will exploit the infrastructure, the level of resources they need, and clearly justify the use of these resources.
2. EVALUATION PANEL: to score the scientific value of the proposal, see the Annex D which includes the scoring criteria. This panel is divided in two subpanels:
    - EXTERNAL: 2 out-of-the-consortium members, Gabriella Zsebeházi (regional climate modeler at the Hungarian Meteorological Service) and Dr. Holger Polhmann (global modeler at Max Planck Institute for Meteorology in Germany, from 25.03.2020, 2nd call). External evaluators must sign the IS-ENES3 Confidentiality Commitment related to Personal Data Protection[23].

[23] https://docs.google.com/document/d/1UaXa8GB033z6mDjqwTpNw4K_Wc1ewZz/edit

- INTERNAL: 2 IS-ENES3 members: Dr. Janette Bessembinder (WP3 co-leader, climate impact) and Dr. Christian Pagé (WP10 co-leader, use of large climate data volumes), researchers at KNMI and CERFACS respectively, with an expertise on the use of climate data in particular for climate impact studies.

3. COORDINATOR & LEAD SCIENTIST: the final decision is made by the IS-ENES3 Project coordinator, Dr. Sylvie Joussaume (CNRS-IPSL), and the Project Lead Scientist, Dr. Bryan Lawrence (UREAD-NCAS).

The secretary provides the administration services in collaboration with the IS-ENES3 Project Manager, Sophie Morellon:

- collects and distributes information, acts as the intermediary for all communications, monitor workflow, and arrange meetings;
- answers the questions of the providers, the evaluation panel members, the potential users, the applicants, and the users (contact person in the web);
- creates training material based on the questions and experience from potential users and helps them in adopting data-near server-side processing workflows;
- disseminates the announcements, presents the service in conferences, and designs and updates the website (text and images);
- assures that the applications are complete and the follow-up forms are filled.

Dr. Maria Moreno (DKRZ) covers this for the first and second call.

A dedicated web page has been designed and deployed at the ENES website[24]. The web page includes all the information on the Analysis Platforms service, including the general documentation and the particular documentation corresponding to each access provider, an online application form, and an animated demonstration of the service (see Fig. 2.6).

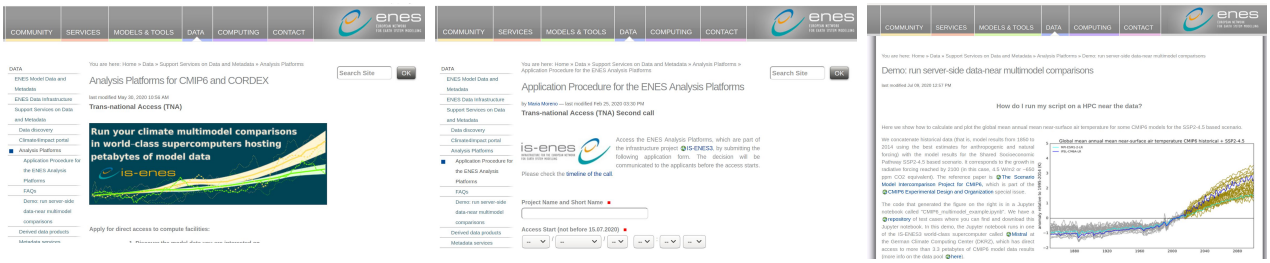



Figure 2.6 TA dedicated web page (left), online application form (center), and demo (right).

[24] https://portal.enes.org/data/data-metadata-service/analysis-platforms

**TA Dissemination**

The service was announced via email (several mailing lists as well as private emailing to collaborators, authors of multimodal comparisons,...) and web (project websites, twitter, linkedin,...). See in Annex E the list of dissemination actions for the 1st and 2nd calls.

**TA Participants**

For the first call (deadline on the 31st of Oct 2019, extended to the 15th of Dec, started the 15th of Jan 2020, ended the 15th of June) we received 10 applications. Many were incomplete, which required administrative support. Besides, 8 of them came from groups with a majority of participants working outside the EU or associated countries and according to the Article 16 of the GA, such groups can only receive 20% of the total access. We encouraged the applicants to find partners working in EU or associated countries. Finally, 2 applications (both with a majority of users working in the EU) passed the administrative, technical, scientific, and final evaluation:

- Project 151219_ie "*High-Resolution Coupled Atmosphere-Ocean-Wave Regional Climate Projections: downscale CMIP6 datasets and ensemble analysis of CMIP6 datasets*" from the Irish Centre for High-End Computing (ICHEC), Ireland. They received 6400 CPU hours at DKRZ. Due to COVID-19 their work suffered delays and they asked for an extension that was granted (see next subsection).
- Project 161219_w "*Evaluating future impacts of extreme weather systems in East Asia and Taiwan with CMIP6 archive*" from the Research Center for Environmental Changes, Academia Sinica, Taiwan, and the Geophysical Institute, University of Bergen, Norway. They received 6000 CPU hours at UKRI JASMIN-CEDA.

For the second call (deadline 31 May 2020, extended to the 10th of June, started in July the 15th and end by Dec 2020), we received 5 applications, all passed the administrative and technical reviews, scientific evaluation, and the final approval by the IS-ENES3 project coordinator and lead scientist. Four groups will access DKRZ and one UKRI. Applicants were notified and only one group did not reply yet. While the selection took place during the RP1, the service will be used during the RP2 and therefore we will summarize the activity in the next report.

**TA Current challenges**

Server-side data-near processing services are a new paradigm for the majority of model data users, which usually (1) download their preferred files from a datacenter and analyse the data on their own premises or, (2) already have access to a HPC center holding the data they need. It has

been challenging to find the user community that is aware of the benefits of server-side data-near server-side processing and does not already have access to HPC centers.

Besides, the users of the first call have exploited a limited amount of resources allocated. When we asked the participants of the first call for the summary of their scientific output (as requested by the EU Commission), the group at UKRI JASMIN-CEDA did not answer yet and the group at DKRZ asked for an extension since their work suffered the inconvenience of the COVID-19 crisis. The extension was granted.

With the aim to increase the number of applications for the second call, the following measures were taken:

1. We offered direct access to the resources for test-activities to accompany potential users in the adoption of the Analysis Platform,
2. We reinforced the efforts in advertising the service (including several contributions and announcements at the international conference EGU2020, see Annex E), and
3. We created a demo[25] including all the steps in an example on how to run a multi-model comparison process at one of the IS-ENES3 HPC (more demos are under development).

However, we only received 5 applications in the second call, a suboptimal number that could be also due to the COVID-19 crisis since the potential applicants might have been apart from their regular activities.

**TA next steps**

We plan webinars with hands-on sessions and asynchronous demos (a blog and video tutorials) on supercomputer literacy (big data software, high performance software, parallelism, and virtualization) in general and, in particular, on how to use our VA and TA services to emphasize the benefits of server-side vs client-side workflows. It will allow us to get feedback to update our services to users' needs and to create more training material.

Besides, and after consultation with our EU project officer, we decided to shift some efforts from the TA to the VA service. The VA service can provide preparatory activities to help users in their posterior use of the Analysis Platforms. This implies some modifications within the Annex 1 of our Grant Agreement regarding the TA service. In the WP7-VA2, the Task2 is centered around the "Comp1" service, which is virtual access, and the Task3 the "Comp2" service, which is our TA service. We think that we would have more applications if we reinforce the Comp1 services, which consists in providing more virtual access to users, so that they get familiar with our service

[25] See Fig. 2.6 and the link in the previous footnote.

and then apply to our TA service (the Comp2). In order to do so, we will need to transfer some PMs from the Comp2 service (Task3) to the Comp1 service (Task2). Therefore, less PMs would be allocated to our TA service, in the objective of attracting more participants via the Comp1 activity.

## 2.4 Support for CF convention and data request (Task4)

### 2.4.1 Introducing the CF Conventions

The conventions for CF (Climate and Forecast) metadata are designed to promote the processing and sharing of files created with the NetCDF API. The CF conventions are increasingly gaining acceptance and have been adopted by a number of projects and groups as a primary standard. The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities.

These conventions provide the foundation for exchange and analysis of all the model data distributed by the IS-ENES3 RI, so every file published and every file downloaded represents a transaction which is dependent on this standard as well as other service components. This usage information will be reported under the core services section of this document.

The CF Conventions are developed and maintained through an international collaboration including partners in the US and European partners who are outside the IS-ENES3 consortium (for example, EUMETSAT is supporting the development of CF in order to exploit the standard for dissemination of satellite data products). The contributions of IS-ENES3 to the development and maintenance of the CF Convention are in two areas: (1) support for the CF Data Model and (2) support for CF Standard Names.

### 2.4.2 CF Data Model

The CF data model 1[26] identifies the fundamental elements (“constructs”) of the CF conventions and shows how they relate to each other, independently of the netCDF encoding. IS-ENES3 funds the maintenance of the cfdm package which implements the CF data model through its internal data structures and so is able to process any CF-compliant dataset. It is not strict about CF-compliance, however, so that partially conformant datasets may be ingested from existing datasets and written to new datasets. This is so that datasets which are partially conformant may

[26] https://doi.org/10.5194/gmd-10-4619-2017

nonetheless be modified in memory. The cfdm package plays a key role in developing and demonstrating the power of the CF Data Model.

The PI for the service is the release of package updates. In the first reporting period there have been two minor releases, 1.7 and 1.8, with 12 and 6 intermediary maintenance releases respectively[27].

### 2.4.3 CF Standard Names

A fundamental requirement for exchange of scientific data is the ability to describe precisely the physical quantities being represented. Within the CF Convention, this requirement is met by the standard name table which contains a register of agreed terms which have been established through a transparent community discussion. IS-ENES3 supports the moderation of the discussion of standard names and the publication of new tables.

The two PIs for the Standard Name service are (1) the publication of new versions of the table: 10 versions have been published (versions 63 to 72) in RP1, and (2) 86 new terms published, taking the total from 4332 to 4418.

### 2.4.4 The CMIP Data Request

The data request of the Coupled Model Intercomparison Project Phase 6 (CMIP6) defines all the quantities from CMIP6 simulations that should be archived. This includes both quantities of general interest needed from most of the CMIP6-endorsed model intercomparison projects (MIPs) and quantities that are more specialized and only of interest to a single endorsed MIP. The complexity of the data request has increased from the early days of model intercomparisons, as has the data volume. In contrast with CMIP5, CMIP6 requires distinct sets of highly tailored variables to be saved from each of the more than 200 experiments. This places new demands on the data request information base and leads to a new requirement for development of software that facilitates automated interrogation of the request and retrieval of its technical specifications.

During the reporting period, the CMIP Data Request resolved 52 issues[28] and released 3 new versions[29] from 1.0.30 to 1.0.32.

[27] See https://pypi.org/project/cfdm/#history.
[28] https://github.com/cmip6dr/ CMIP6_DataRequest_VariableDefinitions/issues and https://github.com/cmip6dr/ Request/issues
[29] https://pypi.org/project/dreqPy/#history

## 2.5 ES-DOC operational support for CMIP6/7

ES-DOC (Earth System Documentation) offers services for metadata search, comparison and creation, following the CIM standard (Common Information Model[30]). It provides an environment to document the modelling workflow. The WGCM and the WIP have tasked ES-DOC with documenting all aspects of CMIP6.

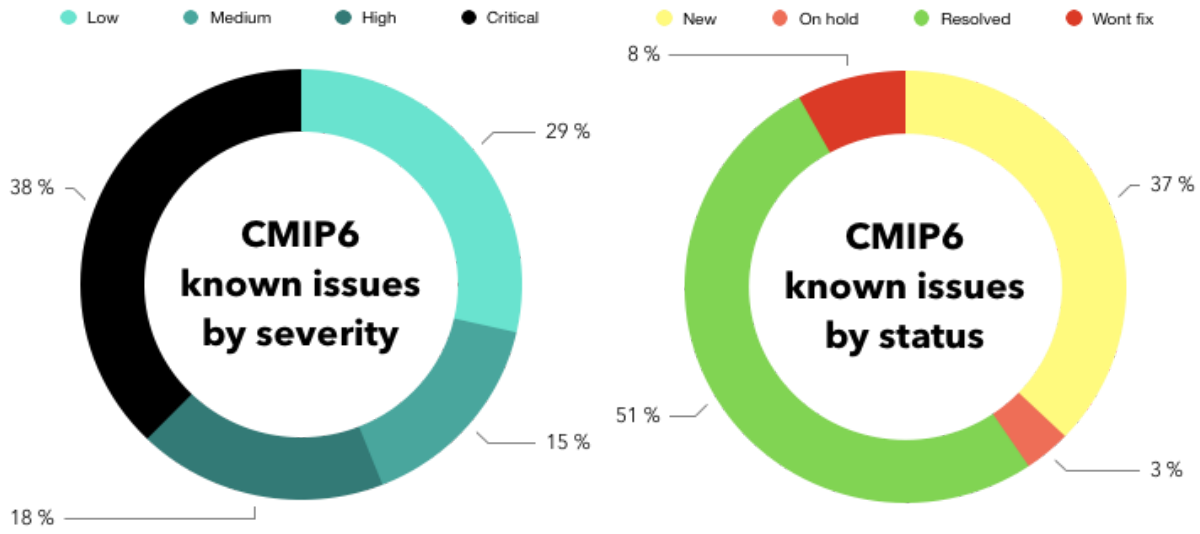



Figure 2.7 CMIP6 known issues severity and status.

ES-DOC brands two main services, which are a Documentation Service for CMIP models, simulations and protocols and an Errata Service for known issues centralization (see Fig. 2.7). In 2018, both ES-DOC services became operational for CMIP6. Through the IS-ENES project, the ES-DOC services are hosted on remote virtual machines provided by WebFaction.

The dataset errata search & publication web-service has had 5705 visits since December 2018. 198 issues have been registered on the web service affecting 89,993 CMIP6 datasets (3% of the CMIP6 archive).

The Documentation search and view web sites have had 21405 visits over the same period. The ES-DOC helpdesk[31] has received 70 questions, requests for support and suggestions since October 2018. These have mainly come from the ES-DOC liaisons at the CMIP6 institutes. The cdf2cim process of the ESGF publisher has generated metadata from ~1000

[30] https://github.com/ES-DOC/esdoc-cim-v2-schema
[31] support@es-doc.org

institute/model/experiment combinations spanning from 2,462,979 publications, from which Simulation and Ensemble CIM documents will be derived.

## 3. Conclusions and Next Steps

As summarized above, key service activities successfully started and supported a large and growing user base. Work started to establish new service offerings especially in the context of flexibly supporting data center near data analysis and processing activities. As outlined in Section 2, currently multiple complementary service offerings are envisioned and the key next step is to reach out to user groups, learning about their requirements and to adapt the service offerings as well as our training and support material. Discussions resulted in activities to address e.g. specifically eastern European user groups not having access to large climate data pools nationally. As mentioned in Section 2.3, the number of TA applications is suboptimal. For the first call it might be attributed to the novelty of the service that takes a while until it reaches the target community. For the second call, the COVID-19 negative effects might have played a role. Before the third call starts (Jan 2021), we will provide webinars and asynchronous demos (a blog and video tutorials) to support TA users in the adoption of our server-side services and get their feedback. Besides, and after consultation with our EU project officer, we decided to shift efforts from the TA to the VA service. The VA service can provide preparatory activities to help users in their posterior use of the Analysis Platforms.

# Annexes

## Annex A: Evolution of the ESGF KPIs: long-term perspective

**ENES CDI ESGF Data download KPIs**

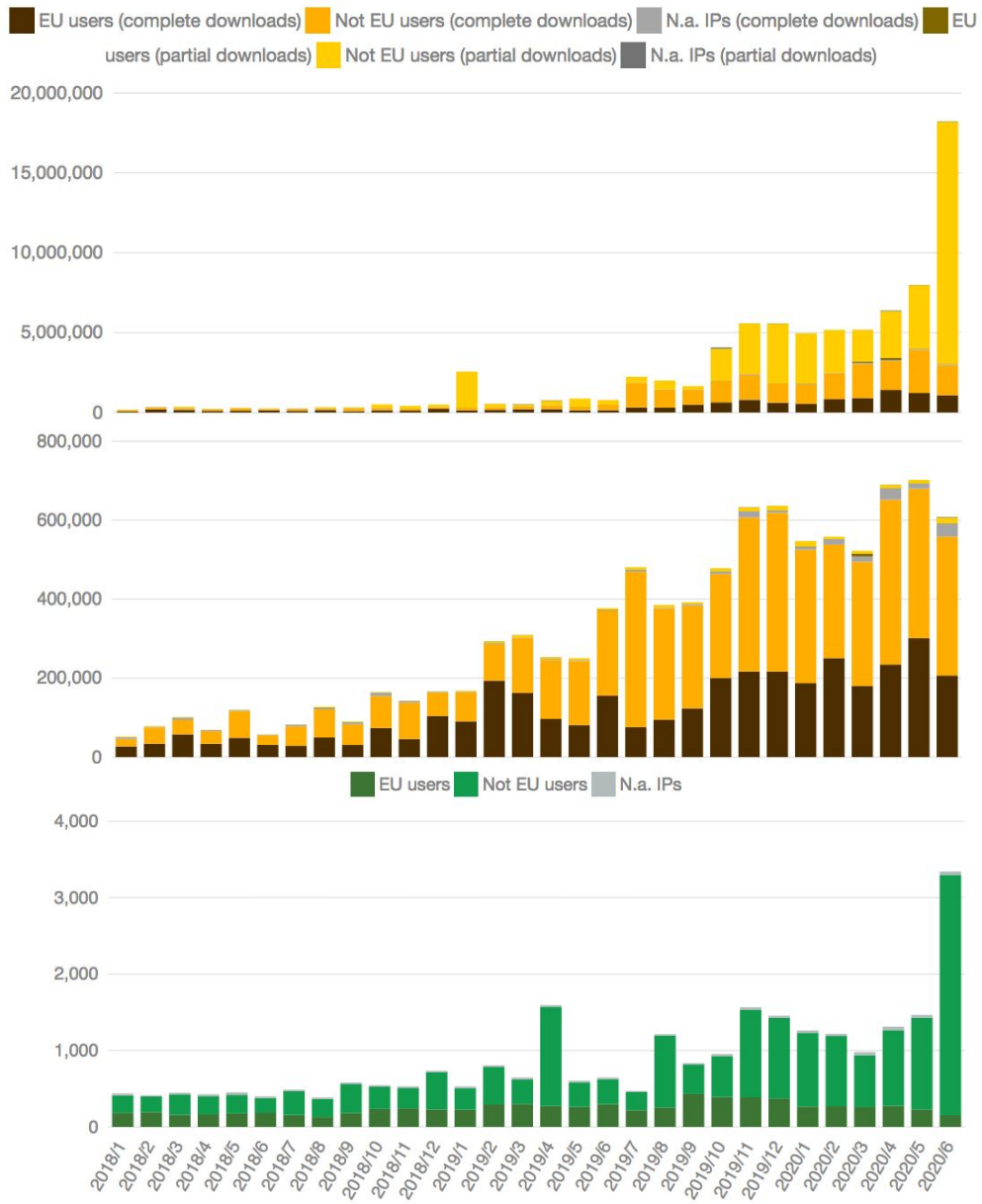



Figure A.1. ESGF data download KPIs (stacked charts): number of downloaded files (up), associated volumes (in GB) (center), and distinct users (bottom) between IS-ENES project phases (2018 was unsupported).

As shown in Fig. A.1, there was a significantly lower activity before IS-ENES3 started supporting CMIP6 data access (first model results came out at mid 2019).

**ESGF User Support Service KPI**

The number of model data user requests and IS-ENES/ENES contributions increased in 2019 with respect to the time between IS-ENES project phases (2018 was not supported) as it is shown in Fig. A.2. "All" stands for answered and unanswered emails (due to the lack of available supporters, some unfiltered announcements, or questions non related to data management issues but model performance, variables, scientific meaning, etc. that are derived to a second level support) to the helpdesk and the percentage is over the total number of threads, that is, including the unanswered emails. When removing the unanswered emails, the ENES/IS-ENES contribution is about 70%.

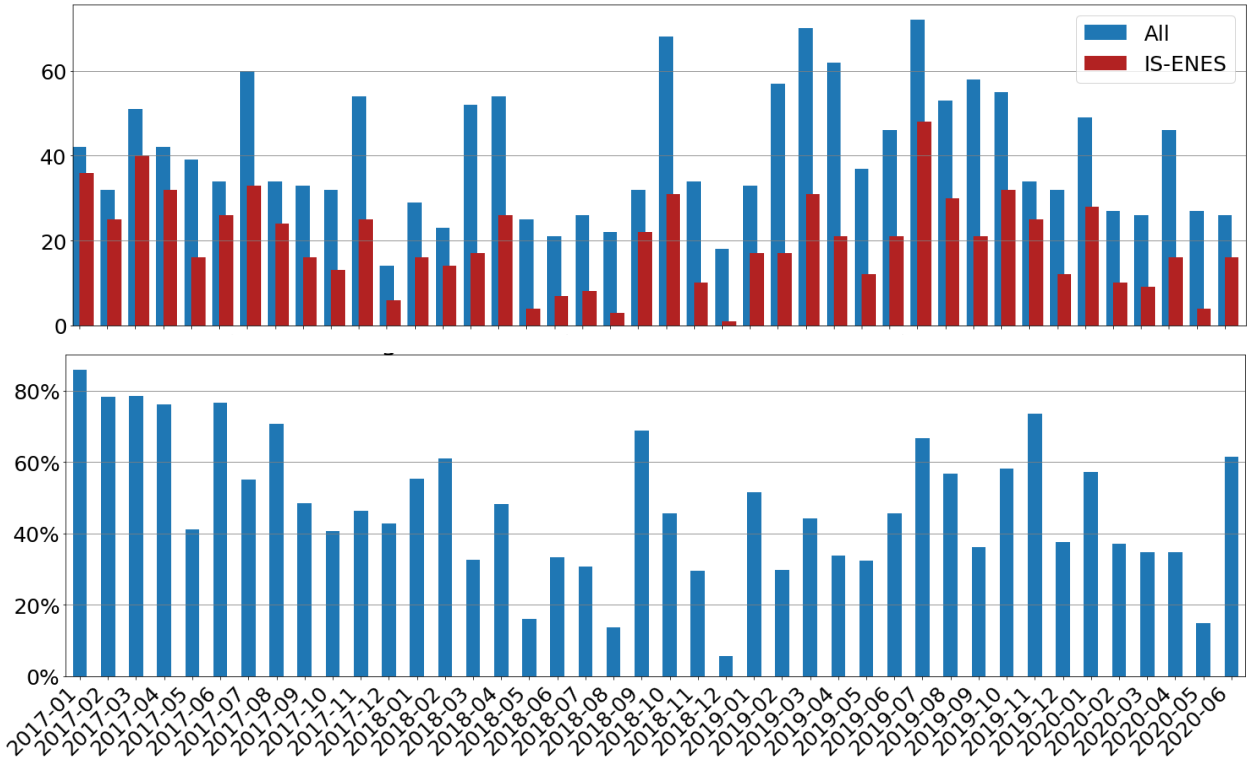



Figure A.2 ESGF user support activity: number of support email threads (up) and the percentage of threads with the contribution of an IS-ENES3 supporter (bottom).

## Annex B: List of ESGF user tutorials and FAQS updates

DKRZ is maintaining the Data Services section of the ENES project website and coordinating the ESGF portals web edition (see Annex C), both linked to the ESGF user tutorials and FAQS documentation. Beside the migration of the complete ESGF user tutorials and FAQS documentation to a github page[32], the following updates been performed during the reporting period:

- Add the Node Status Page https://esgf-node.llnl.gov/status/ in the User tutorial and FAQs
- Add: "if a node is down (usually owing to maintenance), please check the "Show replicas" box in your ESGF search to access data replicas from other nodes." in the User tutorial and FAQs
- Update pyclient link here https://esgf.github.io/esgf-user-support/user_guide.html#own-python-scripts to this documentation https://esgf-pyclient.readthedocs.io/en/latest/
- Update Synda info to here: https://esgf.github.io/esgf-user-support/user_guide.html#synda
- Include: "CMIP5 project data downloads are unrestricted. Please use the -s option to wget scripts to bypass the login prompt" as solution to: 'Error: "HTTP request to OpenID Provider service failed"'
- Fix bad end: "What is the relation between Variable, CF Standard Name and Variable Long Name?" in https://pcmdi.llnl.gov/mips/cmip5/index.html available variables in standard_output.pdf in Output Requirements
- Include the following in the FAQS: 'Where can I find CMIP5 scenarios?': "CMIP6 does not have experiments based on RCP (Representative Concentration Pathway) that were used for CMIP5, because instead scenarios are based on the Shared Socioeconomic Pathways (SSP) instead. You will see these ("ssp...") under the Experiment ID. A web search for "SSP CMIP6" will give you some more information" and the ref paper on ScenarioMIP at www.geosci-model-dev.net/9/3461/ 2016/gmd-9-3461-2016.html the special issue link it is already in the FAQS in "How may I cite esgf and cmip data in my paper"
- In https://esgf.github.io/esgf-user-support/faq.html#there-seems-to-be-missing-data- for-some-models include Errata Service link https://es-doc.github.io/esdoc-errata-client/
- Add that the CMIP6 contact is in the esgf portal page: Search results on esgf-node -> Show Citation -> Data Citation (Landing Page) -> Contacts
- Include in the FAQS and in the mailing list registration page https://esgf.llnl.gov mailing-list.html the info we need from the user when the user send an email to the mailing list:

  *- Your browser and operating system, e.g. Firefox 68.0.2 on Ubuntu 18.04*

[32] https://esgf.github.io/esgf-user-support/index.html#

*- Your OpenID*
*- The complete name of the file(s) or data set(s) you attempt to download*
*- The host node, for instance, esgf-index1.ceda.ac.uk, esgf-node.llnl.gov,...*
*- The steps required to reproduce the problem*
*- The exact error message. Run the script in debug mode to get a detailed error message: $ bash wget-xxx.sh -H -d*

- Regularly check link at https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html and update their link to the tutorial
- Replace CMIP5 by CMIP (general) and/or include links to CMIP6 docs in user guidelines and FAQS
- 'CoG' name appears everywhere. for instance in 'Narrow a CMIP5 data search to just one variable': "CoG has created a way around this problem" it and remove links to https://www.earthsystemcog.org/
- Change the multiple examples of Restful API: http://esgf-node.jpl.nasa.gov/esg-search/search (not working node) by http://esgf-node.llnl.gov/esg-search/search
- Add CMIP5 docs link in the FAQ in Solution 5 for 'I don't find the expected data'
- Add link to esgf errata https://es-doc.github.io/esdoc-errata-client/ in Solution 4 in "I don't find the expected data" and "I have a question or error to report to ESGF"
- Remove deprecated Solution 4 in 'getcert.jar cannot be retrieved': https://esgf.github.io/esgf-user-support/faq.html#esgf-wget
- Update the variables. for instance, in "What is the relation between Variable, CF Standard Name and Variable Long Name?"
  - for CMIP6
    http://clipc-services.ceda.ac.uk/dreq/mipVars.html (variable name)
    http://clipc-services.ceda.ac.uk/dreq/index/miptable.html (frequency)
    https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html (excel name & frequency)
  - for CMIP5
    https://pcmdi.llnl.gov/mips/cmip5/docs/standard_output.pdf?id=99
- Include how to extract modeler info from the file metadata in the FAQ. There were 'Please contact the modeling groups' or 'Contact the data owner' but it is not said how to find that info (for instance, in "I have a question or error to report to ESGF") from the FAQS:
  *For questions related to the models and possibly missing data, please contact the individual modeling centers using the supplied contact information in the file header.*
  *$ ncdump -h <filename> | grep contact*
  *with netCDF4 in python: dataset_name.contact*

## Annex C: List of ESGF portals updates

DKRZ coordinates the effort of reviewing, updating and unifying the look & feel of the six ESGF international portals (esgf-dkrz, esgf-ceda, esgf-ipsl, esgf-liu, esgf-llnl, and esgf-nci). During the reporting period, 21 features have been updated based on new software developments and the supporters' experience in solving issues from the ESGF user support mailing list. Here is the status of the features updates per portal:

| | DKRZ esgf-dkrz | CEDA esgf-ceda | CNRS-IPSL esgf-ipsl | LIU NSI SMHI esgf-liu | LLNL esgf-llnl | NCI esgf-nci |
|---|---|---|---|---|---|---|
| Right sidebar features | | | | | | |
| 1. Technical Support link (remove deprecated Contact Us CoG) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 2. Complete list of Federated Nodes | ✓ | ✓ CoG-CU broken link | ✓ CoG-CU broken link | ✓ | ✓ | ✓ |
| 'Home' tab features | | | | | | |
| 3. Standard content area WITH UPDATED MAP that includes the new nodes: South Korena, Thailand, and Taiwan | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ |
| 4. Link to ESGF | ✓ | ✓ | ✓ | ✓ | ✓ in CoG logo | ✓ |
| 5. Link to IPCC, WMO, WCRP, and/or CMIP | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 6. Links to info on Data Structure, Experim. Design,... smth like CMIP6 | ✓ | ✓ in CMIP6 project | ✓ in CMIP6 project | ✓ in CMIP6 project | ✓ | ✓ CMIP6 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Users Guide or cmip5 info. For CMIP5, please include this sentence: "CMIP5 project data downloads are now unrestricted. Please use the -s option to wget scripts to bypass the login prompt" | | page ✗ CMIP5 sentence | page ✗ CMIP5 sentence | page ✗ CMIP5 sentence | | ✗ CMIP5 sentence |
| 7. Link to enes.org and/or is.enes.org | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 8. Link to es-doc | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| 9. Link to the esgf errata service docs or the portal to look up issues: https://errata.es-doc.org/ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| 10. Link to ESGF account creation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 11. Updated tutorial and FAQs link (the earthsystemcog.org one is deprecated!!) | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| 12. Updated esgf-user support email: esgf-user@lists.llnl.gov | ✓ | ✓ plus CEDA own HelpDesk | ✓ | ✓ | ✓ | ✓ |
| 13. Text on how to contact the modeling groups & the | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |

| | | | | | | |
|---|---|---|---|---|---|---|
| user support mailing list[33] | | | | | | |
| 14. List of Projects with Group Registration | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 15. Link to RESTful API (CoG) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 16. Link to the Data Nodes Status at the LLNL portal | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| 'Contact us' tab features | | | | | | |
| 17. Project contacts | ✓ | link to CEDA contact web | ✓ | ✓ | ✓ | ✓ |
| 18. Technical support | no apply | ✓ | ✓ | | no apply | |
| 'About us' tab features (optional if the info is the same as in 'Home') | | | | | | |
| 19. Partners Sponsors | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ |
| 20. Mission Description | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ |
| 21. Link to the host web | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ |
| Nice to have features | | | | | | |
| announce the new server-side processing capabilities: Analysis Platforms and the Climate4Impact portal | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ |

[33] "For model related questions, please find the modelling group contact info in the header of the netcdf file. For data management issues, please write the following to the esgf user support emailing list:
- Your browser and operating system, e.g. Firefox 68.0.2 on Ubuntu 18.04
- Your OpenID
- The complete name of the file(s) or data set(s) you attempt to download
- The host node, for instance, esgf-index1.ceda.ac.uk, esgf-node.llnl.gov,...
- The steps required to reproduce the problem
- The exact error message. Run the script in debug mode to get a detailed error message: $ bash wget-xxx.sh -H -d"

## Annex D: TA Scientific Evaluation Panel Terms

The panel members that scientifically evaluate the proposals should:

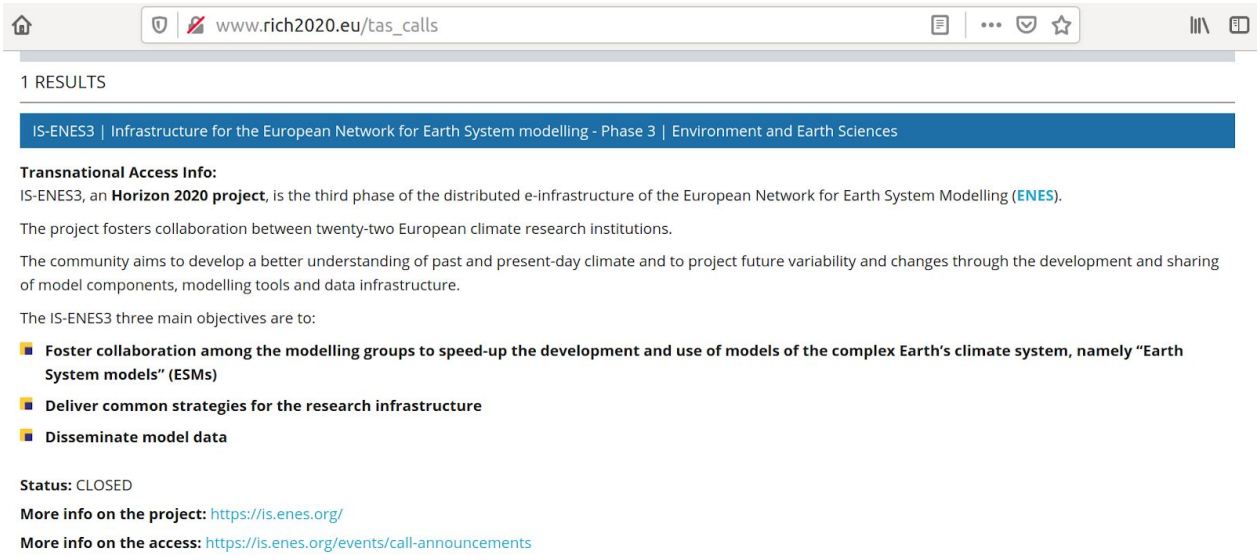
1. Sign the Confidentiality statement for data protection and send it to the secretary (external panel members only).
2. Review and rank the proposals using the scoring guide below.
3. Convene an online meeting to discuss and harmonise the approach to scoring (internal panel members only).
4. Send to the secretary a report, ranking the proposals and summarising the decision process (half-page per application would be enough).

| | Scoring Criteria | Max. Score |
|---|---|---|
| Fit to ENES objectives | • Integration of the Earth's climate system modelling community and the sustainability of its infrastructure.<br>• Common development of models and tools, and the efficient use of HPC.<br>• Exploitation of model data by the Earth system science community, the climate change impact community and the climate service community. | 5 |
| Scientific Excellence | Projects should demonstrate good understanding of scientific issues in their chosen domain, and the ability to develop excellent science through their work-plan. Ideally, reference should be made to the WCRP Grand Challenges. | 5 |
| Project Management | Proposals must set out clear objectives and a well structured timetable, with appropriate milestones, for project execution. Measures to quantify the success of the project in meeting these objectives should be defined (e.g. datasets published, conference proceedings, scientific literature). As IS-ENES3 is not covering any investigator staff costs, projects must demonstrate that they have the resources necessary to exploit infrastructure allocations. | 5 |
| | MAXIMUM TOTAL | 15 |

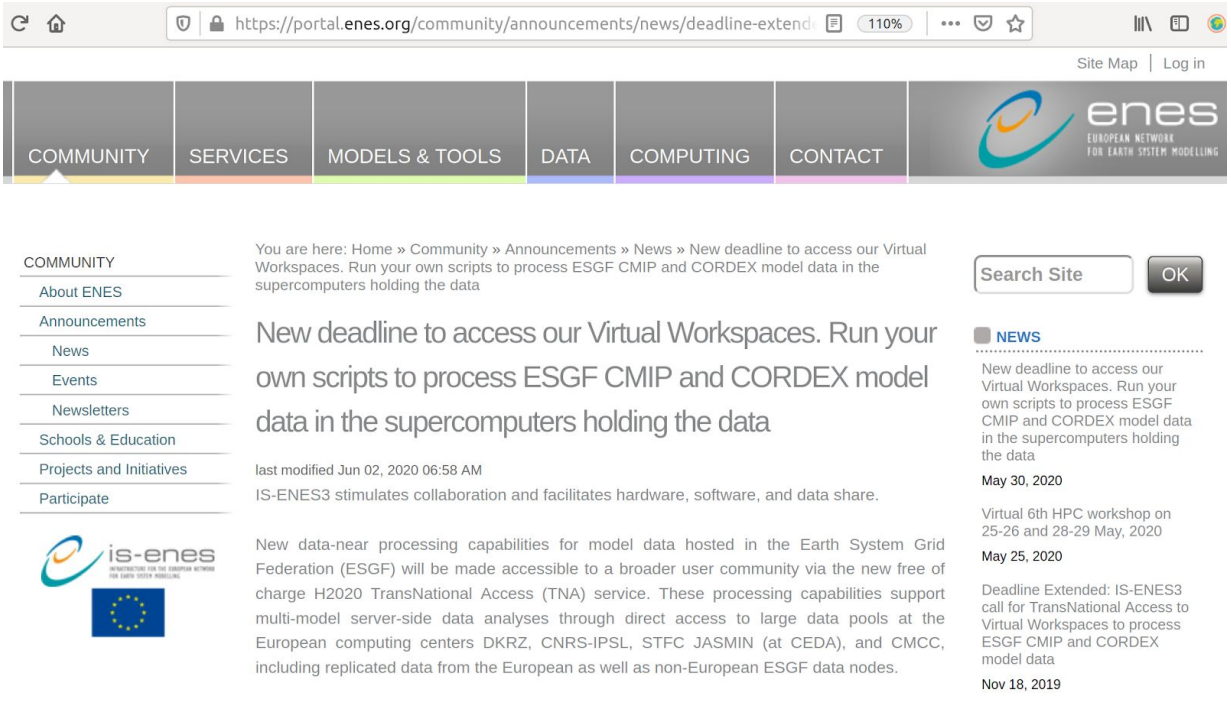
## Annex E: List of dissemination actions

In collaboration with the IS-ENES3 Project Manager, Sophie Morellon, the Project Coordinator, Sylvie Joussaume, and some project members, the TA service has been announced via:

- Webs:
  - Official EU portal rich2020[34]

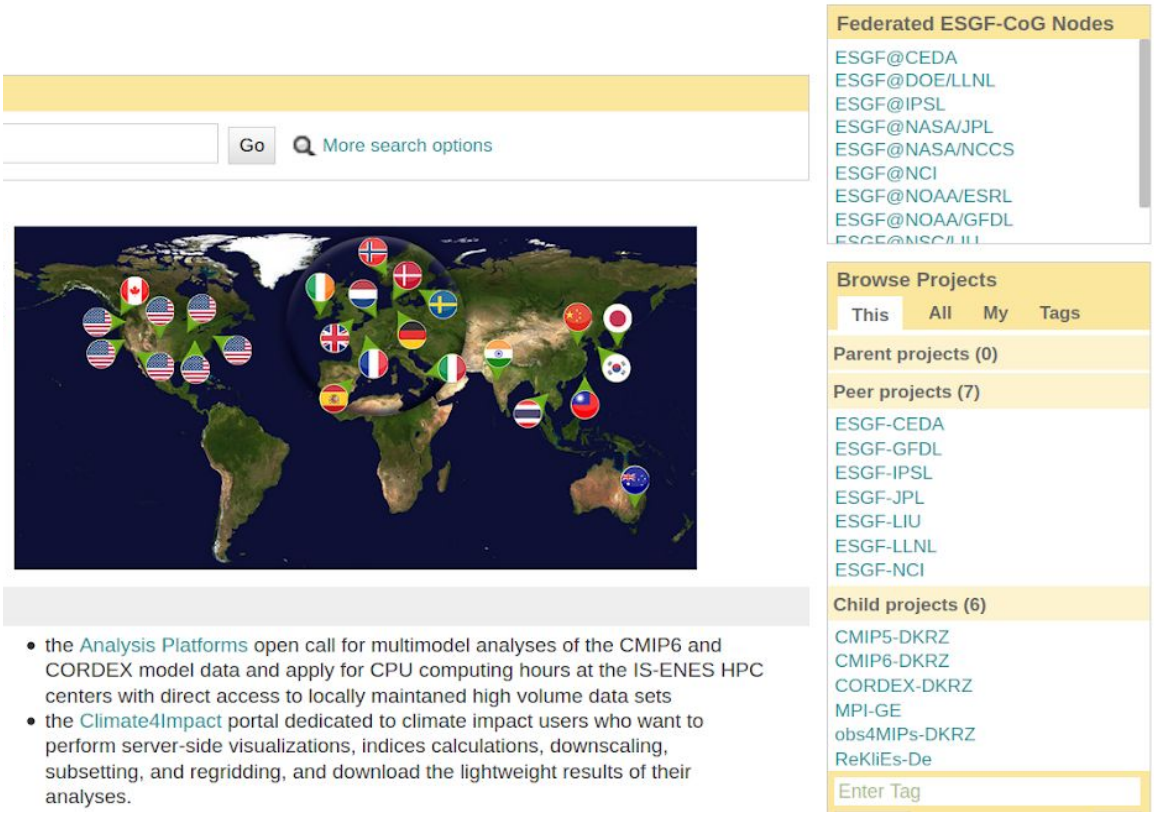



  - Official ENES web https://portal.enes.org/ and IS-ENES3 web https://is.enes.org/



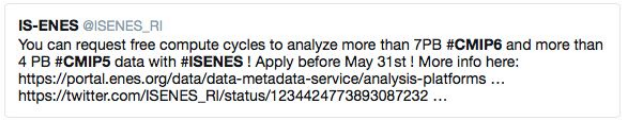

[34] http://www.rich2020.eu/tas_calls

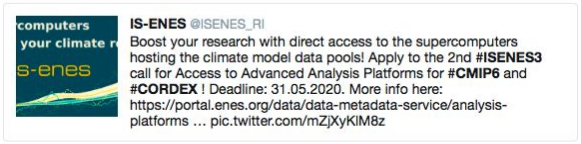
- ESGF European portals, for instance https://esgf-data.dkrz.de/projects/esgf-dkrz/




- Mailing lists:
  - ESGF mailing lists: ESGF users, ESGF developers, ESGF cordex
  - Climate4Impact developers list
  - downscaling list
  - IS-ENES3 list
  - CORDEX communities
  - PRIMAVERA project
  - CRESCENDO project
  - ESMValTool workshop and user list
  - APPLICATE project
  - PMIP list of IPSL
  - EUDAT&PRACE managers and developers at 2019 Summer school
- Newsletters:
  - WCRP newsletter of October 2019: https://us10.campaign-archive.com/
  - Climate Service Partnership newsletter http://www.climate-services.org/
  - IS-ENES3 newsletters

- Twitter:
  - https://www.climateurope.eu/ twitter
  - GEWEX COMMUNITY: https://twitter.com/GEWEX_WCRP/status/
  - IS-ENES3, see some examples below (April 16 2020, March 2 2020, Nov 26 2019, and Nov 13, 2019):



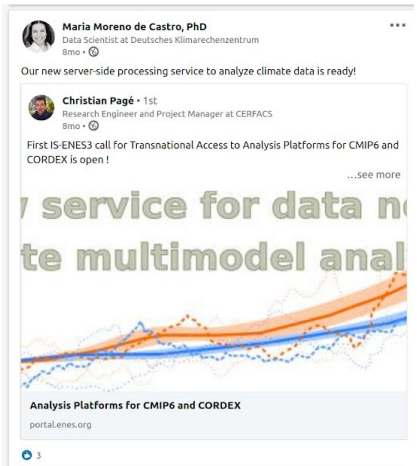









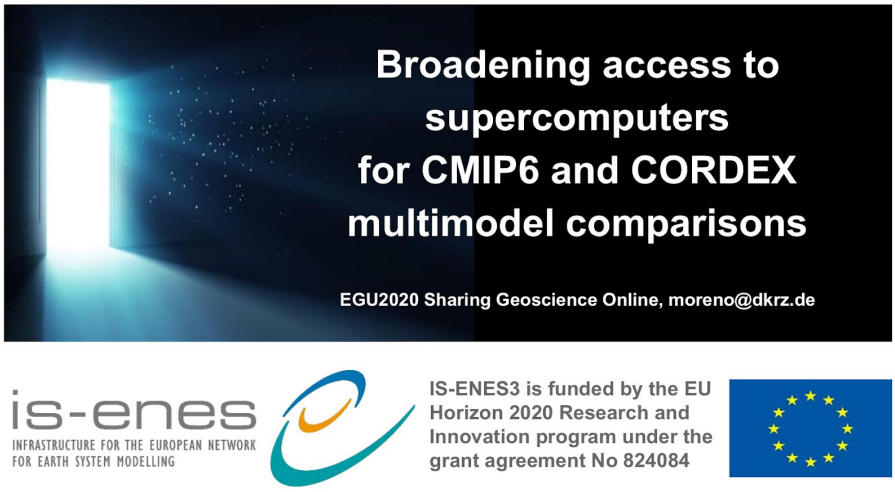

- LinkedIn posts

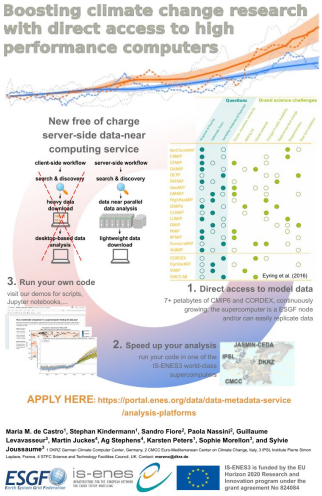



- Groups contacted by email

- Specific outreach to central-east european countries:
  - EUMETNET-Climate Program
  - PannEx Community
- IPCC working groups (via existing connections)
- National Climate service centers
  - GERICS (Germany) www.climate-service-center.de/
  - WASCAL (West Africa) www.wascal.org/
  - SASSCAL (South Africa) www.sasscal.org/
  - Hungarian Meteorological Service
  - Charles University (Prague)
  - Faculty of Physics of the University of Belgrade (Belgrade) are partners
  - ICTP (Italy)
  - CLIMERI-France News: https://climeri-france.fr/

- Conferences:
  - EGU 2020: several announcements, a dedicated talk[35] and a dedicated poster[36]:




- Private emails to existing contacts and authors working in the EU that cited the CMIP6 reference article (Eyring et al 2016).

[35] https://zenodo.org/record/3959252#.Xx_vky3pOgQ
[36] https://zenodo.org/record/3896443#.Xx_wJS3pOgQ