IS-ENES3 Deliverable D7.6

Final KPI and TA report for ENES CDI data services

Reporting period: 01/01/2022 – 31/03/2023

Authors: Stephan Kindermann (DKRZ), Martina Stockhause (DKRZ), Alessandra Nuzzo (CMCC), Martin Juckes, David Hassel (UKRI), Guillaume Levavasseur (CNRS-IPSL), Alessandro Spinuso (KNMI)

Reviewers: Sylvie Joussaume (CNRS-IPSL), Paola Nassisi (CMCC)

Release date: 17/04/2023

ABSTRACT

The ENES Climate data infrastructure provides services for data search, data access and FAIR data management (including support for persistent identifiers and data citation). Associated processing services are established at larger sites and are provided via the Virtual Access and Transnational Access mechanism. Data standards services are provided for the CMIP data request (specifying the variables and controlled vocabularies characterizing the data collections) and the Climate and Forecast Convention (CF). Also dedicated support services are provided with respect to CMIP6 documentation. 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 3 (RP3) report. The evolution of the services is coordinated in close cooperation with WP5/NA4 and WP10/JRA3. Sustainability aspects of the services are agreed and discussed further as part of the sustainability work plan in WP2/NA1.

Revision table			
Version	**Date**	**Name**	**Comments**
0.1	28.02.2023	S. Kindermann, A. Dara	Initial version - ESGF stats pre-filled
0.2	27.03.2023	S.Kindermann, A. Dara, Grigory Nikulin, Guillaume Levavasseur, David Hassel, Alessandro Spinuso, Alessandra Nuzzo	Pre-final version, some contributions missing
0.3	31.03.2023	Paola Nassisi	Review
0.4	03.04.2023	Sylvie Joussaume	Review
0.5	15.04.2023	S. Kindermann et al.	Final version

Dissemination Level	
PU	Public

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

For each of the services provided by the ENES Climate data infrastructure a set of performance indicators is provided. The performance indicators as well as key performance indicators were defined as part of the first report (Deliverable D7.1[1]) and are listed in section 1.1. The service provisioning is based on eleven installations distributed across Europe. More installation specific service details will be provided as part of the IS-ENES3 Reporting Period 3 (RP3) report.

The data delivery related services (see section 2) show a continued high demand in CMIP6 data (especially from non-European users), whereas European users continue to directly rely on the large CMIP6 replica data pools hosted at DKRZ, CNRS-IPSL and UKRI, without the need to rely on the ENES CDI ESGF data delivery services. The service related to the establishment of these data pools is described in section 2.2. Statistics also show a growing need for CORDEX data delivery and access via the ENES CDI ESGF nodes.

The services supporting the FAIR data principles with respect to data identification, citation and long term archival and access are provided in section 2.3, 2.4 and 2.5.

The Climate4Impact portal services were completely upgraded and are characterized in section 2.6.

Statistics for the provisioning of data near compute services are provided in section 3; section 4 summarizes the data standards and documentation related support services.

[1] IS-ENES3 deliverable D7.1 "First KPI and TA report for ENES CDI data services", https://is.enes.org/documents/deliverables/d7-1-first-kpi-and-ta-report-for-enes-cdi-data-services

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

1.1 Overview

In IS-ENES3, installations across Europe join to provide a consistent set of services to the European climate research community, including downstream communities like climate impact research. 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 DKRZ long term archive and the IPCC Data Distribution Centre, DDC), 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 and clusters) hosting the data. Compute service demand is clearly directed towards the VA offering, as this is available without application procedure and available all the time. Only one additional call for TA was issued during the final year of IS-ENES3 (as planned), yet some groups asked for an extension of their existing TA resource allocations and were supported until the end of IS-ENE3 funding. The overall goal of the IS-ENES3 data service activities is to provide operational support to the climate and climate impact research communities and other communities using the model data and tooling provided by IS-ENES3. The VA compute activities are now also closely interlinked with the climate impact community specific service offering via the Climate4Impact portal[2].
The following service activity report provides an update on the performance indicators until the end of 2022 and is structured according to the individual data service areas, reflected in different service tasks: data dissemination, archival and user support (section 2), compute services (section 3) and data standards related services (CF convention and data request as well as ES-DOC in section 4).

1.2 Service statistics and performance indicators

The performance indicators (PIs) and key performance indicators (KPIs) are summarized in the following table and did not change in comparison to the previous reporting period:

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

[2] Climate for Impact (C4I) portal: https://www.climate4impact.eu/c4i-frontend/

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 DOI registered to DataCite
	Number of revisions of citation information published to DataCite
	Number of citation entries added to the service database
Persistent data identification PIs	Number of original and number of replica CMIP6 datasets
	Number of original and number of replica CMIP6 files
DCC 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 for the ENES CDI data services

2. ESGF data dissemination, data archival and Climate4impact services (Task 1)

2.1 ENES CDI ESGF data download KPIs and PIs

The ESGF data download KPIs quantify the monthly *number of files* downloaded from the European ESGF data nodes and the associated *data volume* (with a distinction between complete and partial downloads), as well as the monthly *number of distinct users* successfully performing the downloads.
The KPIs are collected as part of the ESGF Data Statistics service[3] developed and hosted at CMCC. They are summarized in Figure 1.

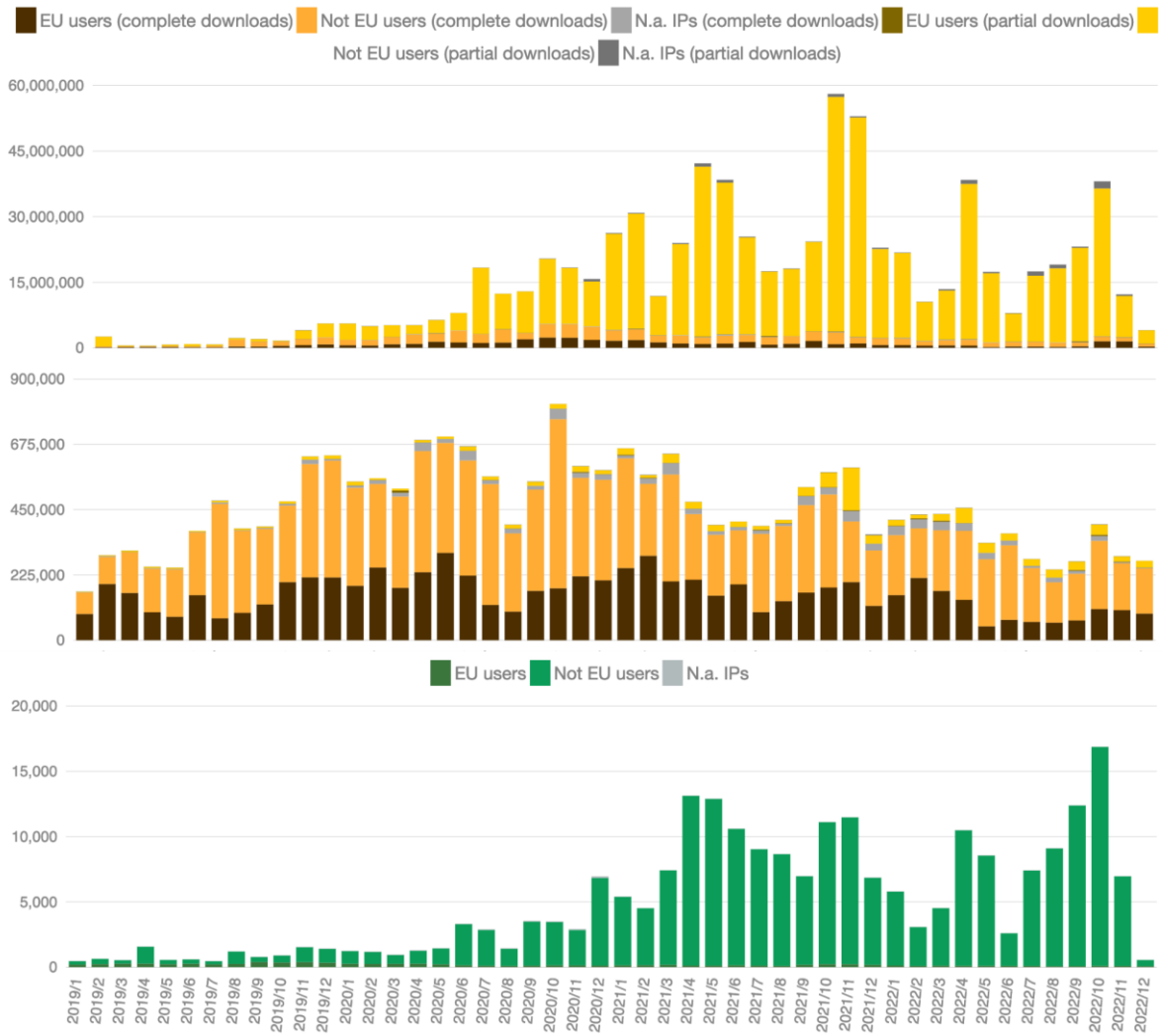


Figure 1: ESGF data download KPIs (stacked charts): number of downloaded files (top), associated data volumes (in GB) (centre), and distinct clients per month[4] (bottom).

[3] http://esgf-ui.cmcc.it/esgf-dashboard-ui/isenes3-kpi.html
[4] 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 other 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.

- **KPI: Number and percentage of emails answered in the user support**

The ESGF user support is mainly handled using the ESGF user support mailing list. We have approximately seven new requests per week and a large portion is answered by IS-ENES3 partners (in particular DKRZ).

Since January 2022, we have received approximately 400 user requests, of which roughly 80% have been successfully resolved. Out of these requests, 142 email threads were related to ENES. However, we have observed a slight decline in the overall number of correspondence compared to previous years, which could be attributed to the increased stability of the nodes.

The most common topic of these requests was troubleshooting data downloads, with some users specifying particular nodes. Among the various projects, CORDEX related issues were addressed more frequently than in the previous years most oftenly related to the additional requirement to request authorization for data access (which is not necessary for CMIP6 data). The number of support email threads for the year 2022 is illustrated in Figure 2, where "All" corresponds to the total number of email threads in the list and "IS-ENES" the portion which is completely answered by IS-ENES members.

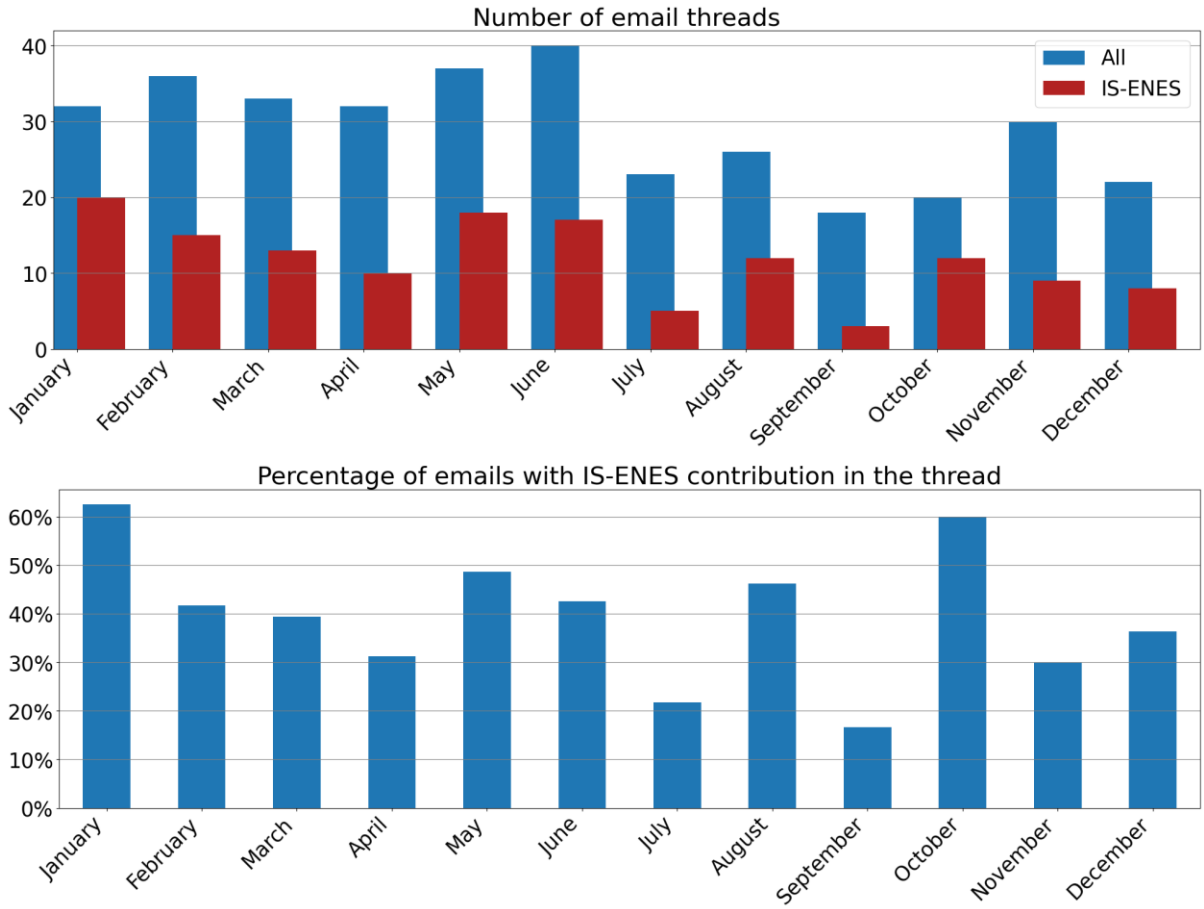


Figure 2: ESGF user support activity: number of support email threads (top) and the percentage of threads with the contribution of an IS-ENES3 supporter (bottom).

- **PI : CORDEX: Number of downloads (EU/no-EU/no geo-located)**

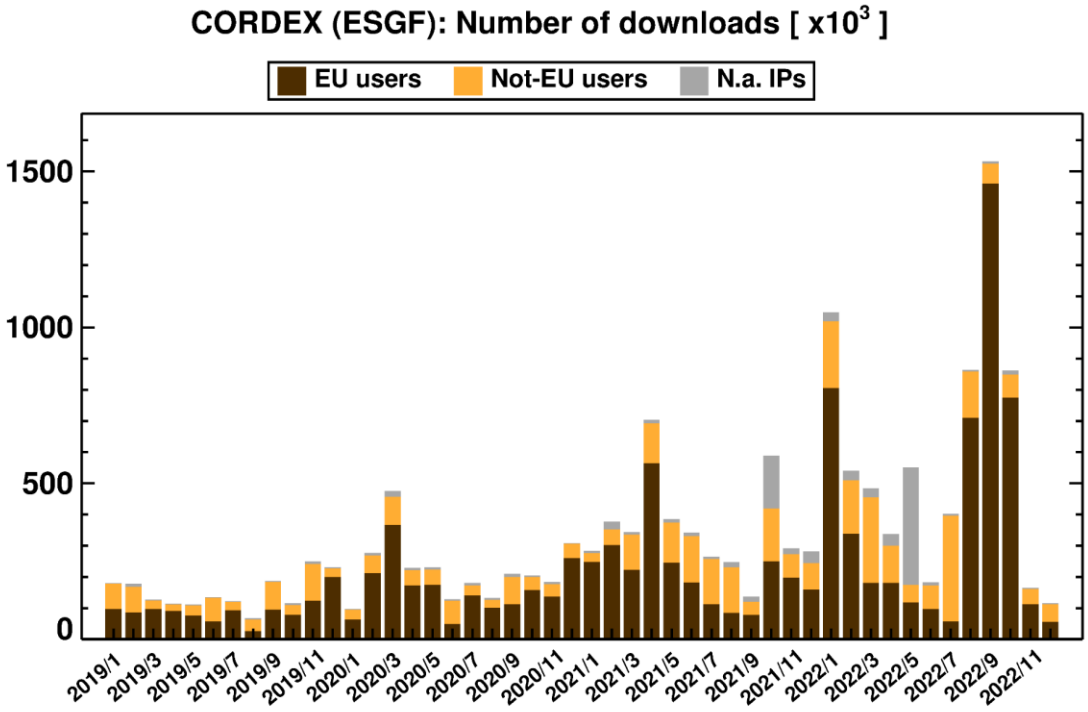


Figure 2: Number of ESGF CORDEX data downloads (file download requests)

- **PI : CORDEX: Downloaded data volume (EU/no-EU/no geo-located)**

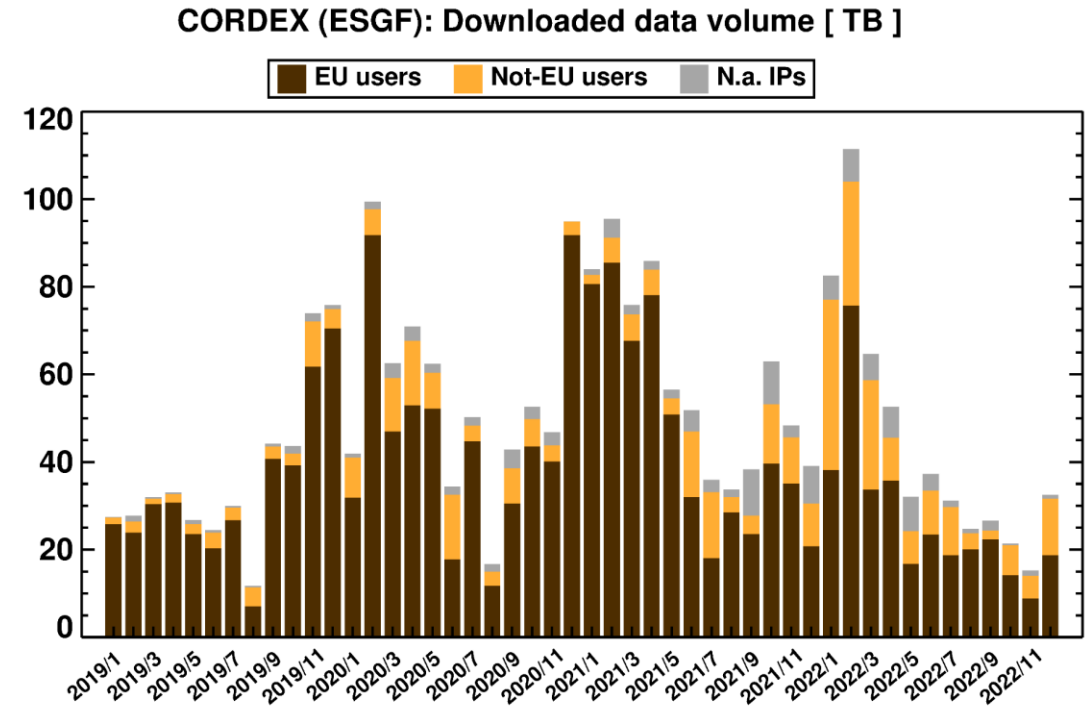


Figure 3: Volume of ESGF CORDEX data downloads

This statistic includes CORDEX datasets for all CORDEX domains. Users from Europe are dominated in the statistics since the CORDEX datasets for different domains are very actively used in many European projects. In contrast, users from other continents are mostly interested in CORDEX datasets for their region only. Additionally, the Euro-CORDEX ensemble (EUR-11) is the largest one among other CORDEX domains and provides many sub-daily datasets that also can explain a large number of downloads by users from Europe.

- **PI : CORDEX: Number of answers to the CORDEX data support mailing list**

The CORDEX data support mailing list (datasupport@cordex.org) is handled by the International Project Office for CORDEX (IPOC) hosted by SMHI. All questions received are sorted first and then forwarded to relevant experts from the CORDEX community. About 30 questions/issues were received during 2022 and all were resolved.

- **Other interesting metrics coming from the ESGF Data Statistics service**

Other interesting data statistics coming from the ESGF Data Statistics service are shown below in Figure 4 and Figure 5 illustrating the distribution by continent of the clients which respectively downloaded CMIP6 and CORDEX data over the European data nodes; Figure 6 shows the top twenty CMIP6 variables downloaded (in GB) from the European data nodes. This statistic is only an indication of the scientific interest in specific variables as the ranking is also strongly influenced by the data organization (variables map to many individual files and file chunking is very different across experiments) and stayed unchanged in comparison to the previous KPI report.

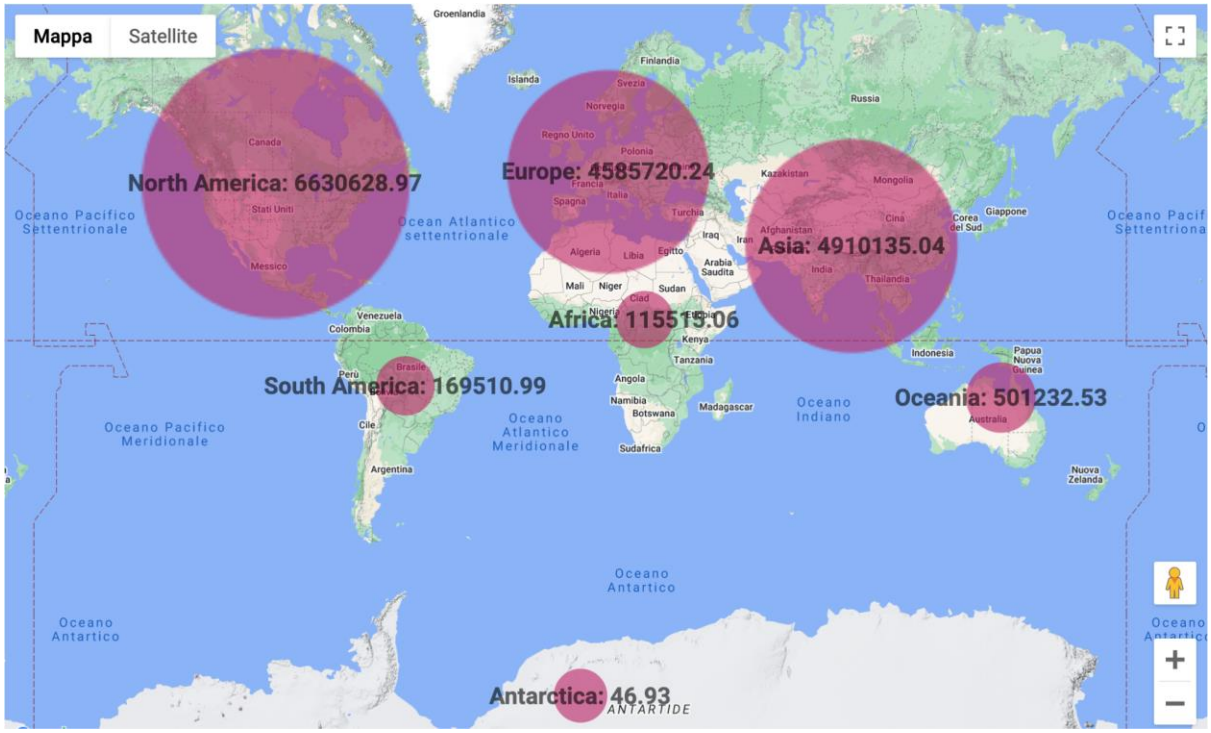


Figure 4: CMIP6 downloaded data volume (in GB) by Continent (from European data nodes)

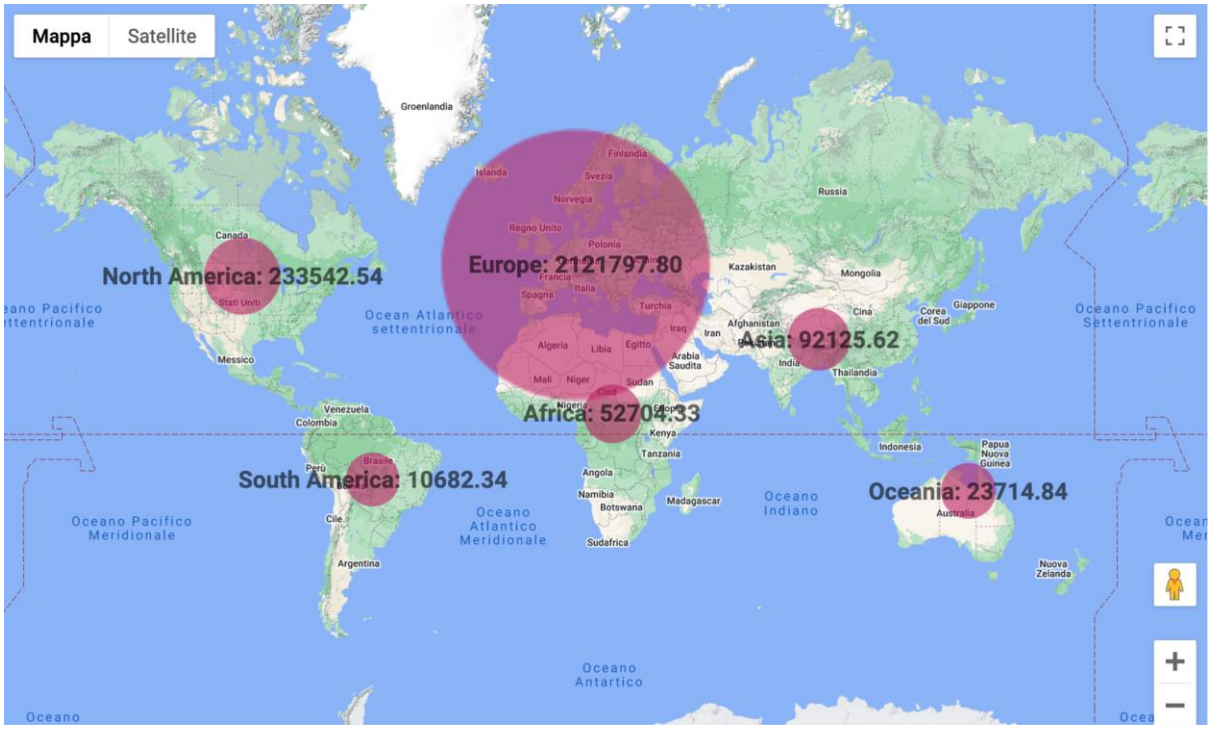


Figure 5: CORDEX downloaded data volume (in GB) by Continent (from European data nodes)

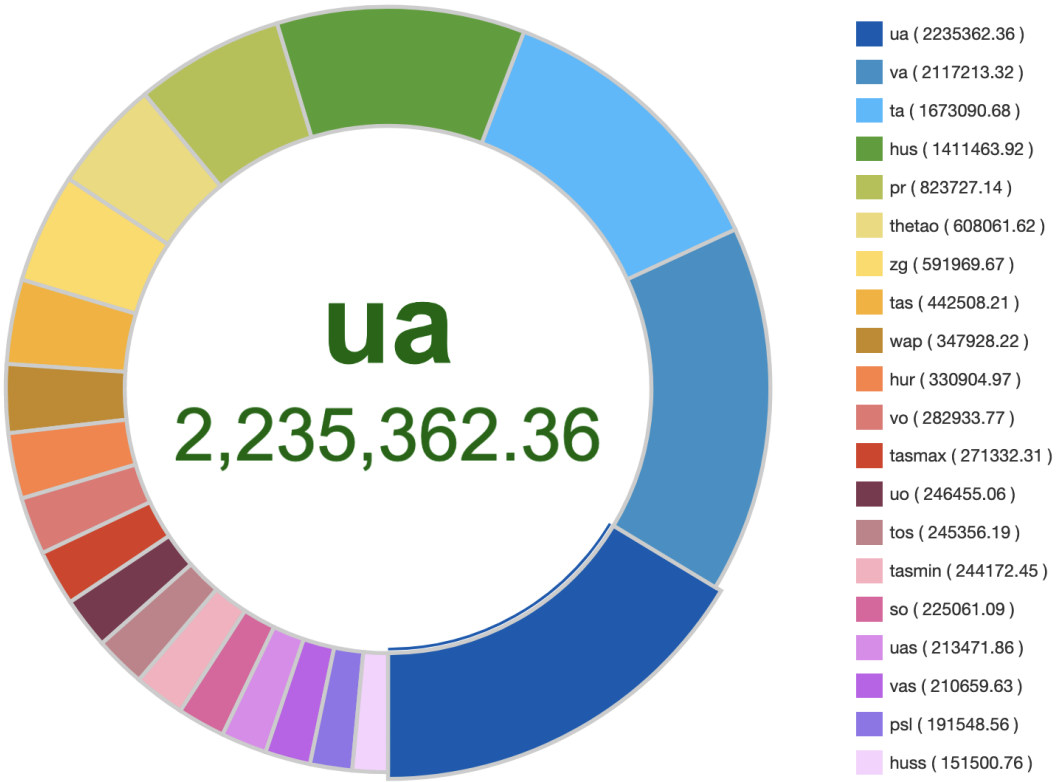


Figure 6: Top twenty CMIP6 downloaded variables (in GB from European data nodes)

The ESGF Data Statistics service also provides general information about the data published over the federation and the projects and data nodes included to date into the environment.
Figure 7 depicts an overview of the available metrics over 25 data nodes, consisting of about 15M published datasets and 1.3 billion downloaded files corresponding to nearly 37 PB of data distributed over 182 countries and belonging to 36 different data projects (with CMIP6 being the largest).

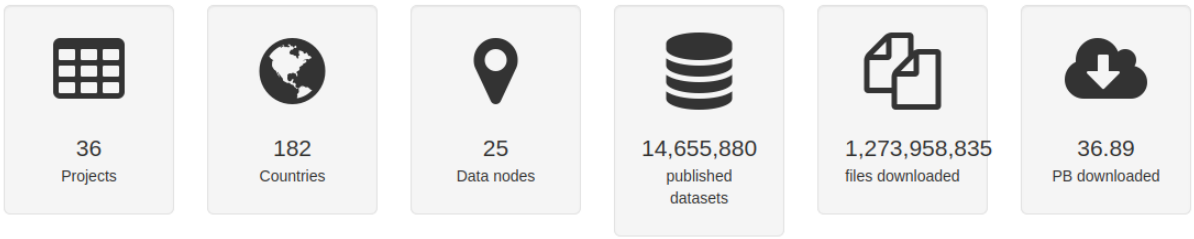


Figure 7: Overview of ESGF

In particular, a total of about 13M datasets and 25 PB of CMIP6 data are available over the whole federation and about 186 thousand datasets and 1.5 PB of CORDEX data (see Figure 8).

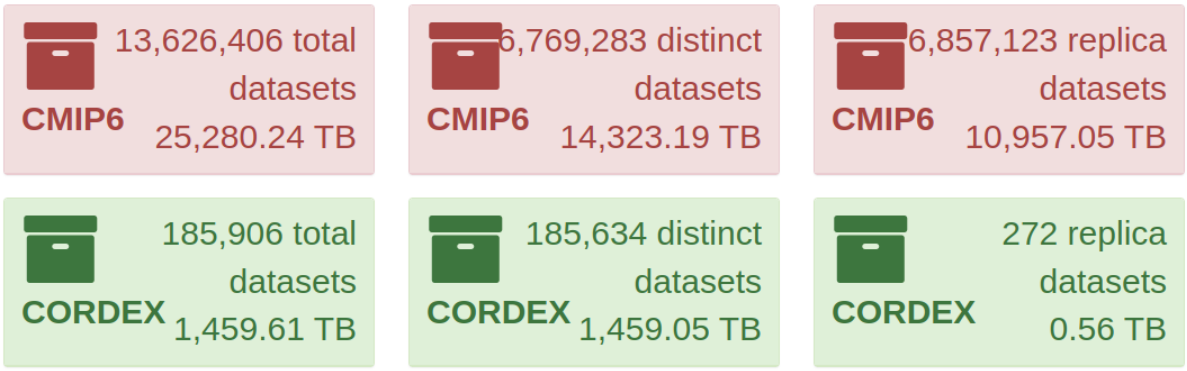


Figure 8: CMIP6 and CORDEX data available over ESGF (as of February 2023).

CMIP6 constitutes the major part of published data in the ESGF. Indeed, the total amount of published data in ESGF is about 35 PB (including replicas) and about 25 PB of this data are from CMIP6. This includes replica, with 14 PB without replica.
On the European nodes, more than 4M datasets and more than 11 PB of CMIP6 data are available. Regarding CORDEX, European nodes provide about 140 thousand datasets summing up to nearly 1 PB of data.

2.2 Replication and Archival PIs

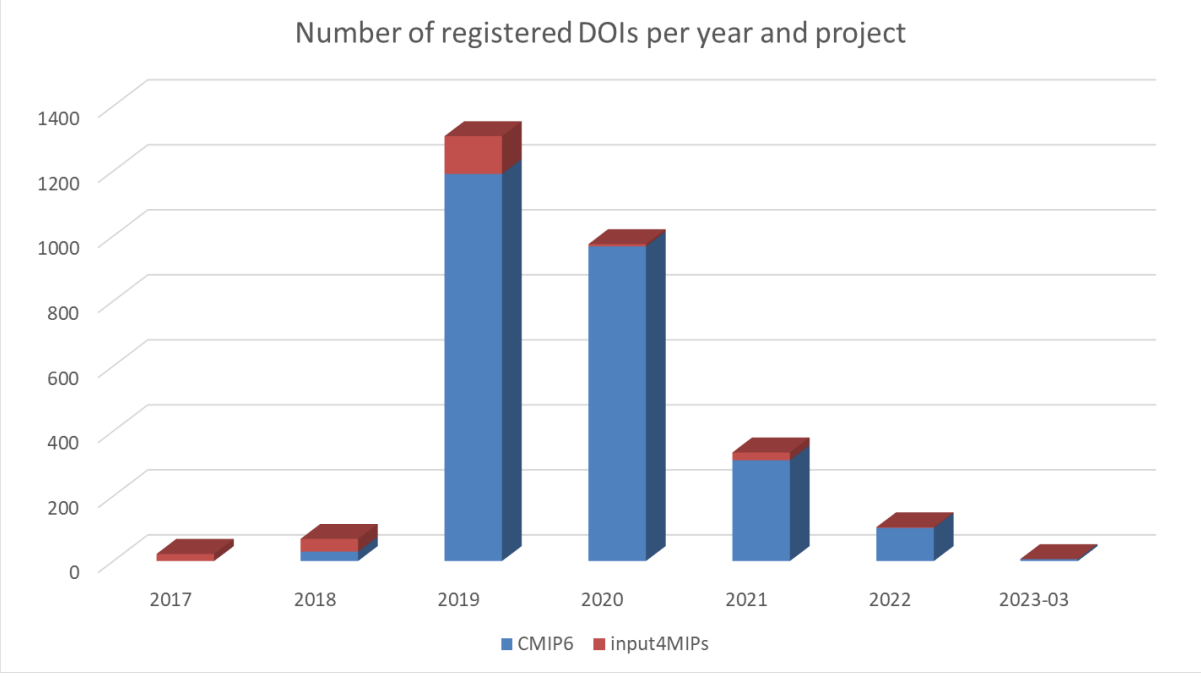
The overall data pool volumes at sites stayed constant, they are essentially limited by the physical storage capacities allocated at sites. The allocated capacities were fully exploited, changes were made with respect to the content (removal and update of new versions of data, deletion of rarely used data and integration of requested data sets).

Institution	Original data (TB)	Replicated data (TB)	Overall volume (TB)
DKRZ	1500	2400	3900
UKRI	1121	1959	3080
CNRS-IPSL	1600	1500	3100

Table 2: CMIP6 data pool volume (original and replicated data collections) at European ESGF sites (March 2023).

2.3 Data citation PIs

By the end of March 2023, more than 2 600 DOIs were registered for the two projects, input4MIPs and CMIP6, and nearly 22 500 Metadata files were registered via DataCite (Figure 9). DOIs are provided on two granularity levels: All data for an experiment run by a model and all data contributed to one MIP by one model and institution. Only for very few datasets in the overall CMIP6 archive no DOI are assigned by now. The CMIP-IPO has taken over the responsibility to contact authors to complete citation information such that DOIs can be assigned. The large number of metadata registrations indicates a large number of metadata changes due to refined information by the authors or adding information through central automated services like adding papers referencing the data to the registered metadata at DataCite.



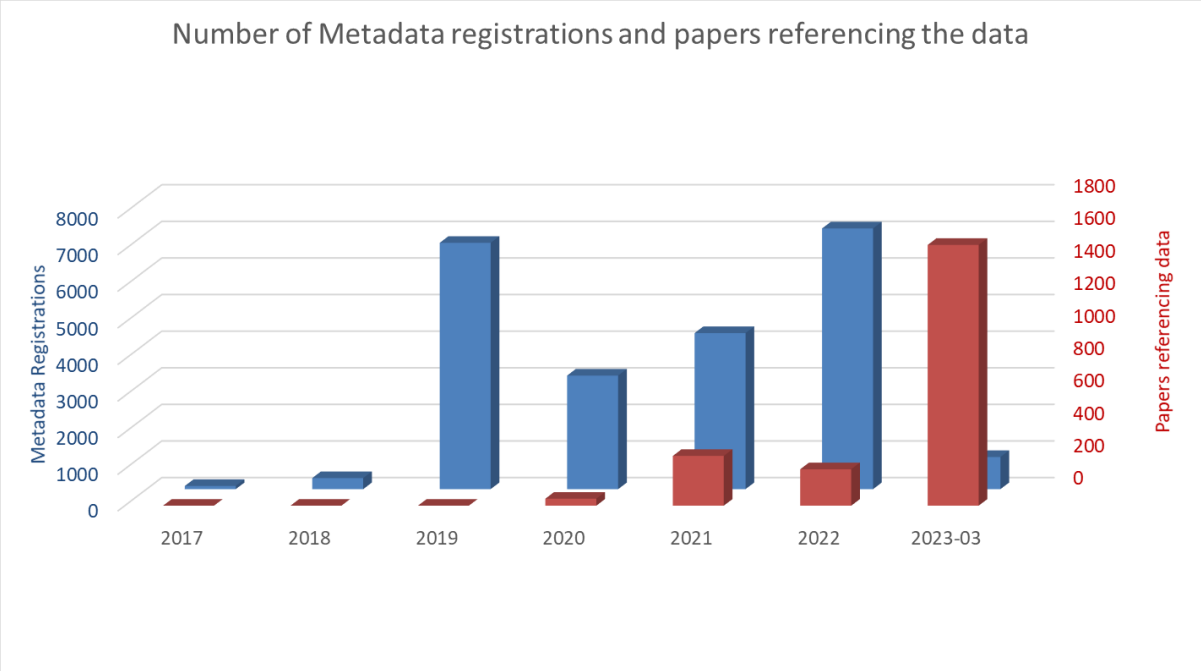


Figure 9: Annual registration of number of DOIs for the projects CMIP6 and input4MIPs (upper part) and number of metadata registrations and number of papers referencing the project data.

The number of papers referencing input4MIPs and CMIP6 data started in 2020 and continues to increase, steeply. Apart from the increase of research based on CMIP6 data, also publishers revise their metadata published to crossref making more of the data references findable through Scholix. In the first quarter of 2023 alone, more than 1 600 paper references were added to the citation metadata adding up to 2 176 paper references in total (Table 3).

A related but nearly finalised activity worth mentioning here is the publication of the DOI-PID relations added to the PID metadata via the OpenAire Scholix Hub. This will improve the cross-linking of DOI and PID level identifiers used in CMIP6 and thus improves the FAIRness of the overall CMIP6 data collection.

Year	CMIP6 DOI Registrations	input4MIPs DOI Registrations	total DOI Registrations	Metadata Registrations	Added paper references (Scholix)
2017	0	22	22	87	0
2018	29	39	68	302	0
2019	1191	116	1307	6728	0
2020	968	7	975	3103	43
2021	310	24	334	4263	306

Year	CMIP6 DOI Registrations	input4MIPs DOI Registrations	total DOI Registrations	Metadata Registrations	Added paper references (Scholix)
2022	102	2	104	7122	224
2023-03	6	0	6	878	1603

Table 3: Annual activities of the citation service (DOI registration, metadata curation and data references in papers added) in the reporting period; numbers according to quality checks documented for Copernicus CDS at https://bit.ly/CMIP6_Citation_Quality.

2.4 Persistent Identification PIs

Persistent identifiers are assigned to data and never removed even if the original data it references is removed. Thus the PI numbers refer to cumulative statistics. As of February 2023, a bit more than 44.1 million PIDs were assigned (~ 6.9 Mio. Dataset PIDs and ~ 37.2 Mio. File PIDs). ~33.8 million of these point to data that is still public, the remaining ~10.3 million point to data that has been unpublished by now (replicas may still be available). Of these ~33.8 million., 80% are assigned at European ESGF nodes (~27.2 Mio.) and 20% (~6.7 Mio.) at non-European ESGF nodes. As PID assignment is directly related to files which are associated to data sets (sets of files per variable over time) the EU/nonEU distribution of PIDs is not directly related to the overall EU/noEU CMIP6 data volume published, it is more related to a smaller dataset / file partitioning used in Europe,

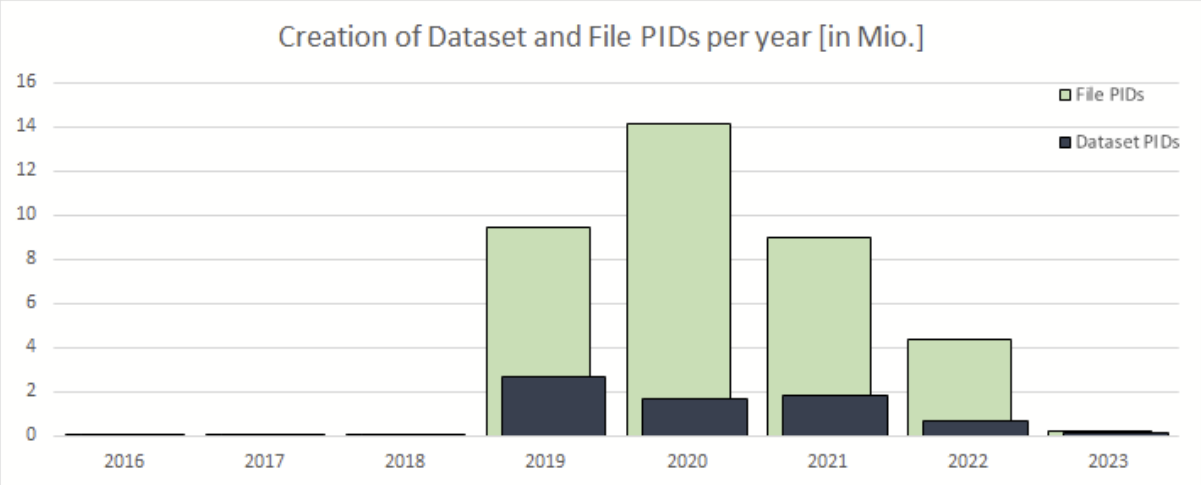


Figure 10: Number of File and Dataset PIDs created per year.

Year	Number of dataset PIDs created	Number of file PIDs created	Total number of PIDs created
2016	8	10	18
2017	295	21	316
2018	34,672	81,843	116,515
2019	2,643,274	9,428,994	12,072,268
2020	1,670,978	14,162,798	15,833,776
2021	1,784,540	8,998,193	10,782,733
2022	693,904	4,364,478	5,058,382
Sum	6,925,093	37,259,813	44,184,906

Table 4. Number of PIDs created (datasets and files)

The decline in PID assignments since 2022 is closely related to the decline of new CMIP6 data publications after the IPCC data cut-off deadline (Jan. 2021) and publication of the IPCC report (August 2021).

2.5 DDC PIs

The data distribution center (DDC) of the Intergovernmental Panel on Climate Change (IPCC) provides an archive for the climate data used in the reports and key figures produced by the IPCC. DKRZ is one of the collaboration partners operating the DDC. As the integration of CMIP6 data in the DDC was completed just recently the DDC statistics still mainly refer to CMIP5 data stored. The downloads of IPCC-related CMIP datasets from all phases continues to decrease in 2022. The reason is the decline in CMIP5 data downloads whereas CMIP5 data is the largest share of datasets in the DDC. The long-term archival of the CMIP6 data subset underpinning the AR6 WGI has not been finalised, yet. The status of the archival process is documented in the milestone document M7.3 and the archival will be continued during 2023 without IS-ENES funding. The reason for the delay is the installation of the new tape long term archival backend during 2022 and associated archival service downtimes during 2022. The average monthly downloads in 2022 were 39 000 datasets/month and 21 TBytes/month via the DDC portal and for CMIP5 also through ESGF (Figure 11). The ESGF download share is more than 90% of the dataset downloads (thus only 10% were accessed via the dedicated DDC portal.

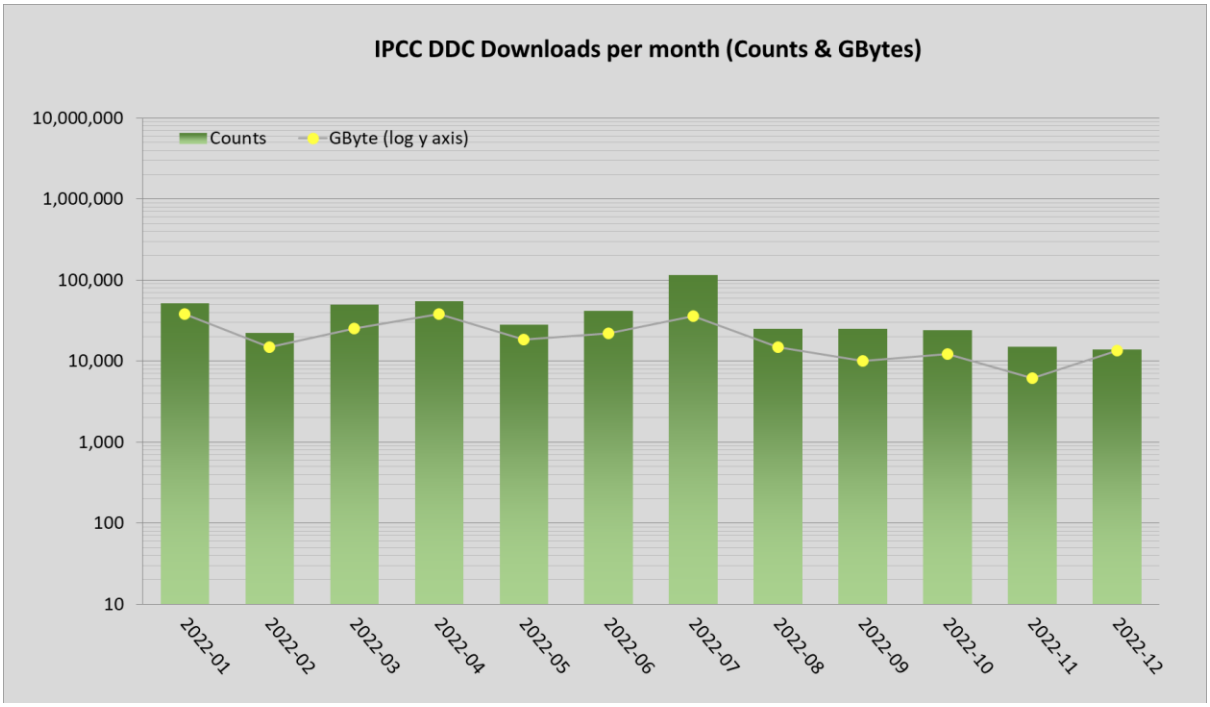


Figure 11: Monthly data download volume and dataset counts from the IPCC DDC at DKRZ in the reporting period. Numbers include direct downloads from the DDC as well as downloads through ESGF.

The continental distribution of users downloading DDC datasets is dominated by Asian users as usual (Figure 12). European users make up 17 % of the total downloads.

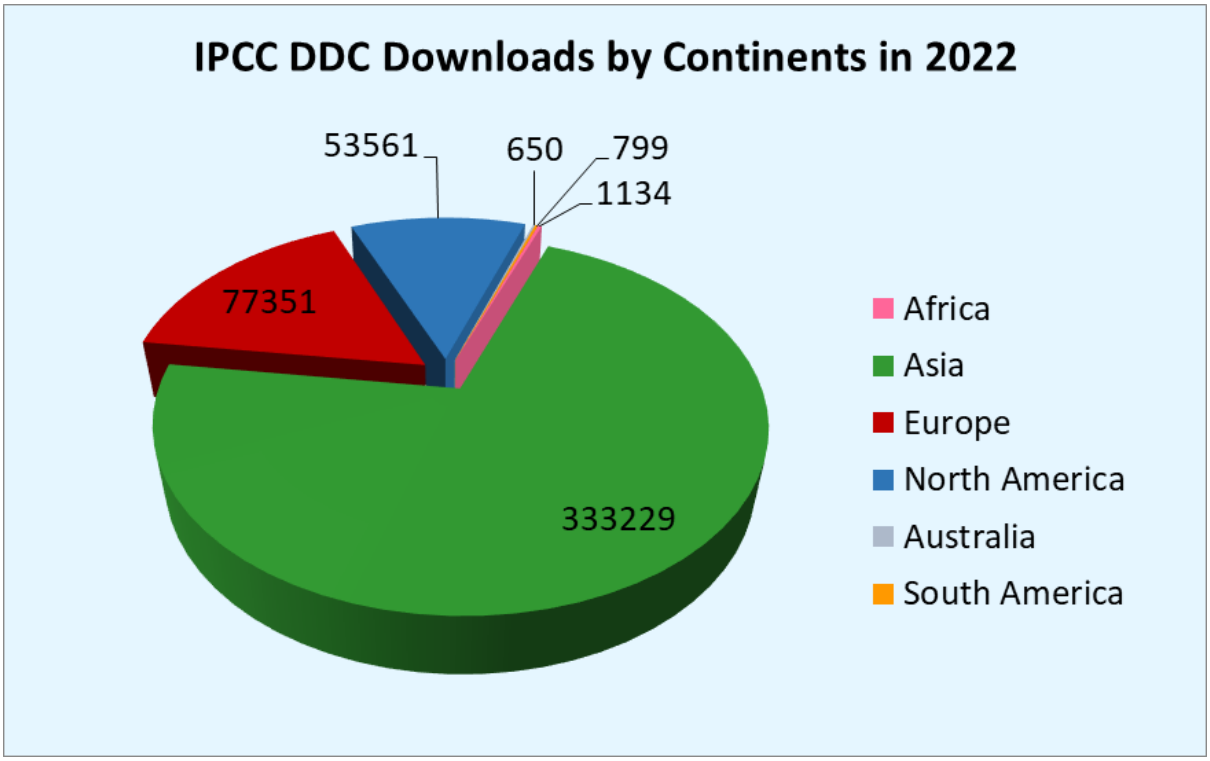


Figure 12: DDC dataset download counts via ESGF and the DDC portal per continent of user's residence during the year 2022.

A detailed analysis of DDC downloads in 2022 is available in the DDC Annual Report (Stockhause, 2023[5]).
The long-term archival of the CMIP6 input data subset and the AR6 intermediate datasets in the IPCC DDC AR6 Reference Data Archive (which is not work planned as part of IS-ENES3) is nearly finalised with few remaining CMIP6 dataset archivals and the DOI registrations after the completion of the quality assurance.

2.6 Climate4Impact v2 KPIs

Climate4Impact (C4I) is a portal that enhances the discovery of climate research data and enables experimentation within impact analysis-ready workspaces. In the last phase of the project we opened to the public Climate4Impact v2[6], which is the new official release of the system. The transition phase required the old site to be dismissed, making the collection of its metric less relevant given the extended downtime of the service, thereby of little use for this reporting period. The new site was open on an unofficial URL for test and evaluation purposes. In February 2023, after having exposed and demonstrated the portal in training events, we officially opened the new service to the public via the official internet address.

For this version of the portal, we have refined and updated the collection of KPIs to better represent the new services offered and goals achieved by users. Besides the most traditional web based KPIs. We now take into account metrics associated with new underlying components aimed at the provision of JupyterLab workspaces, as well as the execution of data-staging and processing workflows, which are both managed by the SWIRRL component, see D7.2 and D10.3. Below the list of refined metrics, divided by the applied gathering methodology, which we describe in the following section.

Web-analytics metrics

1. KPI Visits (Total Visits, new Visits, Returning Visits)
2. KPI Visitors (Unique Visitors, Unique Returning Visitors)
3. KPI Number of local-data download (MetaLink file downloads, each file includes multiple data-requests).
4. KPI Number of access requests to the personal workspaces (Jupyter Notebooks)

Provenance metrics

5. KPI Number of requests for new workspaces
6. KPI Number of remote workflows executed by users, their type and success rate.
7. KPI Total number of Files staged to workspaces
8. KPI Number of Unique Files staged to workspaces (respect to the actual ESGF node)

System metrics

9. KPI Number of users registered to the service
10. KPI Number of active workspaces (Jupyter Notebooks in use)

[5] Stockhause, Martina. (2023). Report 2022 of the DDC at DKRZ. Zenodo. https://doi.org/10.5281/zenodo.7554078
[6] https://www.climate4impact.eu

Metrics for C4I v2 in operation in 2023 January - February - March	
Web Analytics	
Total Visits (2023)	1900
Returning Visits	940
Unique Visitors	(stats available on monthly basis, avg of 50 unique visitors per month)
Unique Returning Visitors	(stats available on monthly basis, avg of 35 unique visitors per month)
Number of local-data download (MetaLink files with multiple data requests)	241
Number of access requests to the personal workspaces	138
Provenance Metrics	
Number of requests for new workspaces	34
Number of remote workflows executed by users.	98 data-staging (91% success rate), 18 wps (77% success rate)
Total number of Files staged to workspaces	1104
Number of Unique Files staged to workspaces	949
System Metrics	
Total number of users registered to the service	91

Table 5: Metrics for C4I v2 in operation in 2023

The KPIs have been recently refined and almost entirely implemented. Metrics have been consistently collected only for a few months of operations, thereby we cover a limited time scale. We provide below an overview of the stats for those metrics which cover the first part of 2023. However it has to be taken into account that many registered users have started to experiment with the advanced workspace feature already in 2022, when the portal was in development and in *alpha* phase, thereby before putting in place this KPI strategy. Hence, the numbers reported in the table are indicative and underestimated.

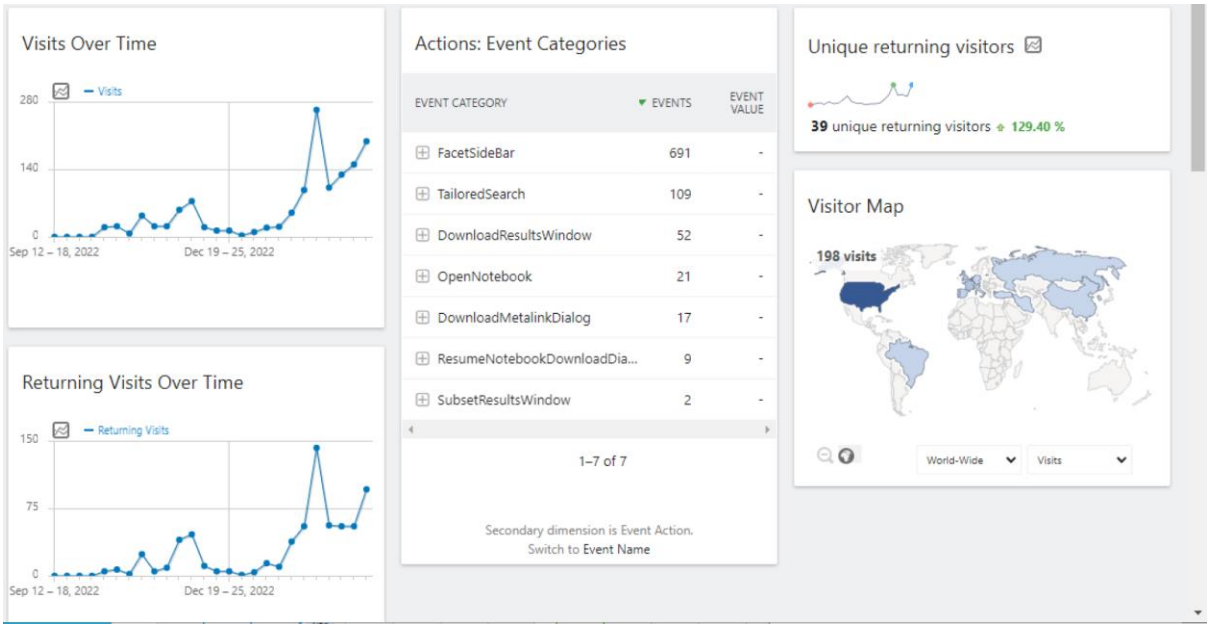


Figure 13: Dashboard of Matomo for a weekly-based web analytics metrics showing, on the left, progress from the first activation of the metrics in September 2022, when the portal was still in *alpha* release. The peak in the visits is related to a week of training events. Matomo Action and Events Categories allow a very high granularity of analysis of the user interaction and goals. The map on the right shows indications of people's linguistic origin, depending on their browser settings.

2.6.1 Analytics sources and tooling

For the collection of usage and access data, we have implemented complementary systems to produce different KPI/PIs. Metrics 1 to 4 are collected via a modern web analytics tool, Matomo[7] (see Figure 14). The service is GDPR compliant and KNMI hosts one instance to gather data for C4I, as well as other KNMI web-facing services. The following metrics are obtained by querying and analyzing the provenance records produced when offering computational services to the users. The records are generated by the SWIRRL system and archived in the provenance graph database. Figure 14 shows two of the many PROV Templates[8] in use. We foresee in future, possibilities to extract more metrics on the use of the computational environments, for instance, gaining insight on the popularity of software modules, datasets or subsetting regions, in combination with particular research interests. Metrics 9, KPI Number of users registered to the service, is obtained taking into account users who were granted a registered profile, after they had submitted their request, as explained on the portal[9]. Although users sign in in SSO using the current ESGF IdeA login system, thereby beyond the control of C4I the registration procedure is necessary to validate the user, which will then be able to access computational resources. This type of access requires mitigation measures to prevent cyber-security risks and optimisation of costs. The latter are sustained in-kind by the institute. We can fetch the number of registered users and active workspaces (workspace expires after two weeks of no use), by interrogating our internal user management

[7] https://matomo.org/
[8] http://dx.doi.org/10.1109/TSE.2017.2659745
[9] https://www.climate4impact.eu/c4i-frontend/register

database, which is decoupled from the provenance data, to meet the GDPR regulation in terms of handling of personal data management. Last metric, concerning active workspaces, when implemented, will interrogate directly our K8S cluster, that allocates and culls these resources, on demand or based on our workspace expiration policy.

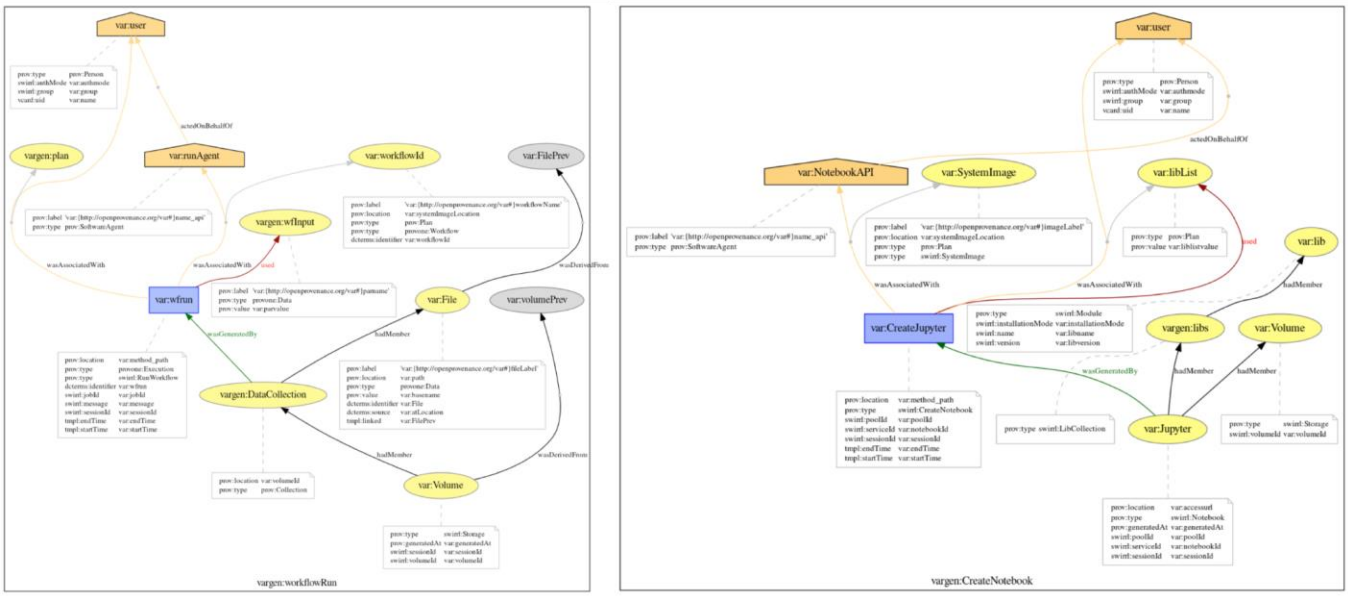


Figure 14: View on the provenance templates gathering information about the creation of a C4I notebook workspace and the associated data-staging/subsetting workflows. The images show the provenance relationships between Data and Software entities (Yellow), Activities (Blue), and Agents (Orange) involved in using the data and the computational environment, according to the notations of the PROV W3C standard, which have been extended by contextual metadata (in the white boxes).

2.6.2 Climate4Impact User support

We provided continuative support for operations of the Climate4Impact v2, reacting to downtimes and supporting users facing difficulties in accessing its functionalities. We are open to consider larger involvement into training and education activities, as indicated by interest already expressed in this direction.

To facilitate interaction and guidance with users, we have implemented a user-feedback page, and published help material explaining how to search and download data locally, as well as perform data staging and subsetting workflows to proceed with custom analysis in the workspace. For the latter, we also published a collection of sample analysis notebooks that can be executed in C4I v2. Additional guidance material is in the process of being finalised by experts and it will be delivered beyond the project's end. This is fundamental to provide specific background information on the scientific use cases and workflows leading to an effective climate-impact analysis, as recommended by the external reviewers. Finally, we have activated a dedicated email address where users can ask questions directly to the C4I development team, and get support on any particular issue.

3 Compute services

Compute services are provided by 4 installations at DKRZ (Germany), UKRI (UK), CNRS-IPSL (France) and CMCC (Italy). They are separated into light weight low resource usage access services provided under the virtual access mechanism as well as access to compute platforms which are provided under the transnational access mechanism (and which involve an application review procedure, which is explained in detail in the Deliverable D7.1). Two installations (DKRZ and UKRI) continued to provide the transnational access possibility.

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

CMCC

The tables below provide some information about the exploitation of the Virtual Access services hosted at CMCC. The first table shows access metrics to the ESGF Data Statistics service web interface, which reports data usage and publication statistics about the ESGF federation. The second table, instead, provides information about the CMCC Analytics Hub web portal, the web access point to the CMCC Analytics Hub, which provides a user-friendly data analytics environment to support scientists in their daily research activities.

ESGF Data Statistics service web interface

From 01/01/2022 to 31/12/2022

Total users	Sessions	Page view	Countries
446	1000	2991	47, mostly from United States, Italy, UK, France, Germany, Australia, China

CMCC AnalyticsHub web portal

From 01/01/2022 to 31/12/2022

Total users	Sessions	Page view	Countries
897	1150	2076	117, mostly from United States, Japan, India, Bangladesh, China, Mexico, France, Spain

DKRZ

DKRZ virtual access platform (login nodes, Jupyter hub portal, CMIP data pool).
From 01/01/2022 to 31/12/2022

Registered Users	Active Users (> 0 node hours)	Node hours used	Countries
84	19	> 4800	**Western Europe:** UK, Norway, Germany, France, Netherlands, Italy, Sweden, Spain, Portugal, Ireland **Eastern Europe:** Czech Republic, Poland **Other:** India

The VA reporting concentrates on the specific accounts allocated to the IS-ENES dedicated VA service project at DKRZ. Many other projects exist allowing users to access the platform but they are not included in the metrics above.

CNRS-IPSL

The generic VA compute access at CNRS-IPSL has been maintained during this final reporting period and fully described in the "Final Release of the ENES CDI software stack document" - D10.5. During this period, the number of active users of the IPSL Compute and Data centre slightly decreased (Figure 15) as the number of CPU.hours. In the same time, the number of jobs launched at the computing centre increased by 50%. This can be explained by a sharp increase in the use of the JupyterHub deployed into production during this period. Each Jupyter instance is counted as a job but not part as CPU.hours consumption because both VA are built over different infrastructures.

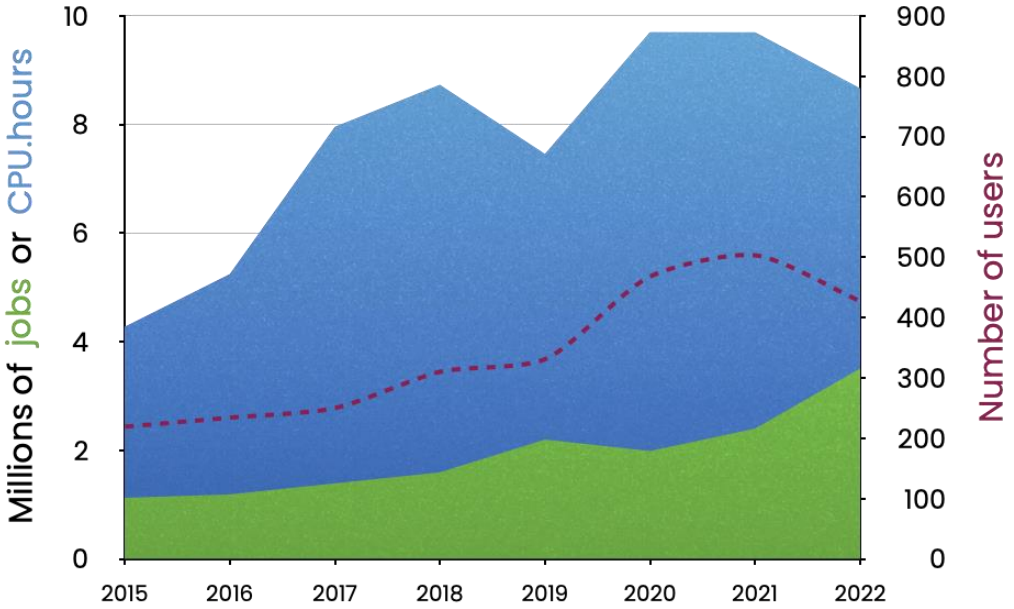


Figure 15. Compute service usage at CNRS-IPSL in terms of jobs and CPU-hours

The WPS deployment has been enforced during the first quarter of 2023 as planned in the RP2. It will be enforced beyond what was planned again during summer 2023 with new physical machines hosting the service with a higher number of CPUs. The WPS will be opened to CNRS-IPSL computing center users and extended VA access to a broader EU community (not only in the Copernicus context).

UKRI

The UKRI service provided through the JASMIN data analysis cluster supports access to the CMIP6 archive and a wide range of additional datasets including re-analysis data, satellite observations, and many smaller datasets. Many of the datasets, including CMIP6, are open access to encourage wide and flexible usage, so statistics specific to each dataset are not available. There are 2,612 active JASMIN users[10]. The service provision in millions of core hours over the last 4 years has been: 31, 35, 49, and 45 (for 2019 to 2022).

3.2 Compute service: Virtual workspaces (Transnational Access - TA, Task3)

The current report includes the last application period (December 2021 to December 2022) that followed three transnational access allocation periods spanning from July 2020 to January 2021, February to July 2021, and July 2021 to January 2022. The table below provides a summary of the number of successful applications received during each of these periods and their corresponding allocation to the TA service providers, DKRZ and UKRI.

successful applications	allocation period	assigned to DKRZ	assigned to UKRI/Jasmin
2	Jan 2020-June 2020	1	1
5	July 2020- January 2021	3	2
6	February 2021 - July 2021	3 (1 switched to VA service)	2
3	July 2021 - December 2022	2	1
5	December 2021 - December 2022	5	0

Table 5: Successful TA applications

The current report pertains to the 5th allocation period, which spanned from December 2021 to December 2022. Among the 5 successful applications, only one project utilized a significant amount of HPC resources. However, several projects from the previous allocation periods requested extensions and were able to utilize additional resources. The PIs of these projects

[10] The users come from: University (1445), NERC Research Centres (349), governmental (UK Met Office and National Laboratories such as STFC: 651), Commercial (46), others (121).

have also provided us with final reports on their projects. Consequently, we have also included these projects from the previous allocation periods in the list below.

DKRZ

5th allocation period (December 2021-December 2022) :

- **(06_12_22_es)** "Post-processing of Climate Models Outputs": switched to the VA service. The project PI noted a high degree of importance of our service for their project. Project outcomes have been submitted to a scientific journal. Further extension has been requested. The project was carried out by a private company from Spain in close collaboration with a university.
- **(14_12_22_ro)** "Evaluation of the radiation budget in climate models": extension requested from this project from Romania.
- **(20_12_22_hu_1)** "Analysis of the Mediterranean cyclone characteristics during the 21st century under different SSP scenarios": no activity. The PI from Hungary took part in project **20_12_22_hu_3.**
- **(20_12_22_hu_2)** "Effects of melting of the arctic and greenland on midlatitude circulation": no activity. The PI from Hungary took part in project **20_12_22_hu_3.**
- **(20_12_22_hu_3)** "Raising public awareness on the attribution of the local climate extreme events to global climate change - Climate Attribution in Hungary": HPC resources were used to analyse CMIP6, CORDEX and ERA5 datasets. 1 PhD student and 4 MSc students were involved in the project. Project outcomes were presented at 11 conferences.

4th allocation period (July 2021 - January 2022):

- **(23_05_21_no)** "Eval weakening Overturning": the project from Norway, under the CMIP6 framework, has been successful with one paper submitted to a journal. However no substantial HPC resources were used.
- **(28_05_21_nl)** "Accessing Climate Change impact on the Hydrological Cycle using the eWaterCycle Platform for Open Hydrology": the project results from Netherlands were used in a MSc thesis and will be published as a part of the PhD thesis of the same student.

3rd allocation period (February 2021 - July 2021):

- **(201020_es)** "Modelant": the project from Spain used a substantial amount of resources. The project is still ongoing, and produces scientifically valuable outcomes.
- **(301020_no)** "NORTH, A NORTHern perspective on CMIP6 climate model variability": the PI from Norway noted the usefulness of the IS-ENES infrastructure at DKRZ for their project that mostly leveraged ESMValTool. One article has been submitted and the second one is to be submitted in March 2023.
- **(301020_1_no)** "DecNorth, Multi-model assessment of decadal climate predictability in the North Atlantic": the PI from Norway reported that "The results of this work will be published as the second paper necessary for the PI to obtain their Ph.D. title. The infrastructure offered and the support team's good quality have been essential to making

the analyses faster and more reliable. Otherwise, access to these datasets would take longer and fewer models would be used in relation to the ones currently analyzed."

- **(311020_cz)** "MCS LOVE CCS, Multimodel Climate Simulations - LOcalization, Validation and Evaluation of Climate Change Signal": supported, switched to ECAS (VA) service. The PI from Czech Republic stated a crucial role of the TNA service for their project. The project results were published in one article and presented at two conferences.

UKRI JASMIN

4th allocation period (July 2021 - January 2022):

- **(30_03_21_il)** "Stratospheric Nudging And Predictable Surface Impacts (SNAPSI)": CEDA have been providing science support to the project from Israel to enable the implementation of metadata standards based on the CMIP6 data standards. The outcomes resulted in one publication with several more to be published in 2023.

3rd allocation period (February 2021 - July 2021):

- **(271020_uk)** "PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment (PRIMAVERA)": according to the PI from UK, "This access has been invaluable and so far has contributed to nine papers that have been published in peer reviewed journals that acknowledge the IS-ENES3 funding <...>. A further four papers have also been produced but were unable to acknowledge the funding. Further papers are still in preparation."

4. Data standards and documentation

4.1 Support for CF convention and data request (Task 4)

CF Data Model and CF standard names

UKRI supports the development and maintenance of the CF standard name table and other CF controlled vocabularies. Updates to the standard name table were published in March 2022[11] and February 2023[12], together accounting for the addition of 80 new terms. By the end of March 2023, a further 39 names have been agreed for publication in the next version of the standard name table.

CMIP Data Request

UKRI-STFC supports the CMIP Data Request. There have been 16 releases of the Data Request: 7 in 2019, 8 in 2020 and 1 in 2022, of which 5 are pre-releases (beta), 6 are patches, and 5 full releases.

[11] http://cfconventions.org/Data/cf-standard-names/79/build/cf-standard-name-table.html
[12] http://cfconventions.org/Data/cf-standard-names/80/build/cf-standard-name-table.html

During the project period 48 issues were resolved; much of the data request development took place before the project started, with other 400 issues resolved in total for the CMIP6 cycle, in addition to numerous email discussions not recorded in the KPI.

4.2 ES-DOC operational support for CMIP6 (Task 5)

The Errata documentation service is currently between versions, as the first iteration remains the main endpoint providing the service, the alpha phase for version 2.0 is currently underway, this is a validation phase designed to confirm the evolution's goals were reached and remove any, if ever, unwanted behaviour or bugs. This next iteration brings more openness to the community suggestions of errata entries, while the errata officers will retain their privileges and also get moderation duties for incoming errata propositions. A notification sub-system is also included in the new version to provide updates regarding moderation status. The work has been developed and deployed in a test environment during RP3. The errata v2.0 system will be released in a production environment effectively replacing the previous version in June 2023.

Breakdown of the errata entries:
The ES-DOC Errata service indexes **421** CMIP6 issues, **8** CORDEX issues and **1** input4MIPS issue, bringing the total to **430** issues currently in the database. 38 of which were created and updated during RP3.
In terms of severities and status, the following pie charts show how these issues breakdown with a focus on CMIP6 and CORDEX related issues:

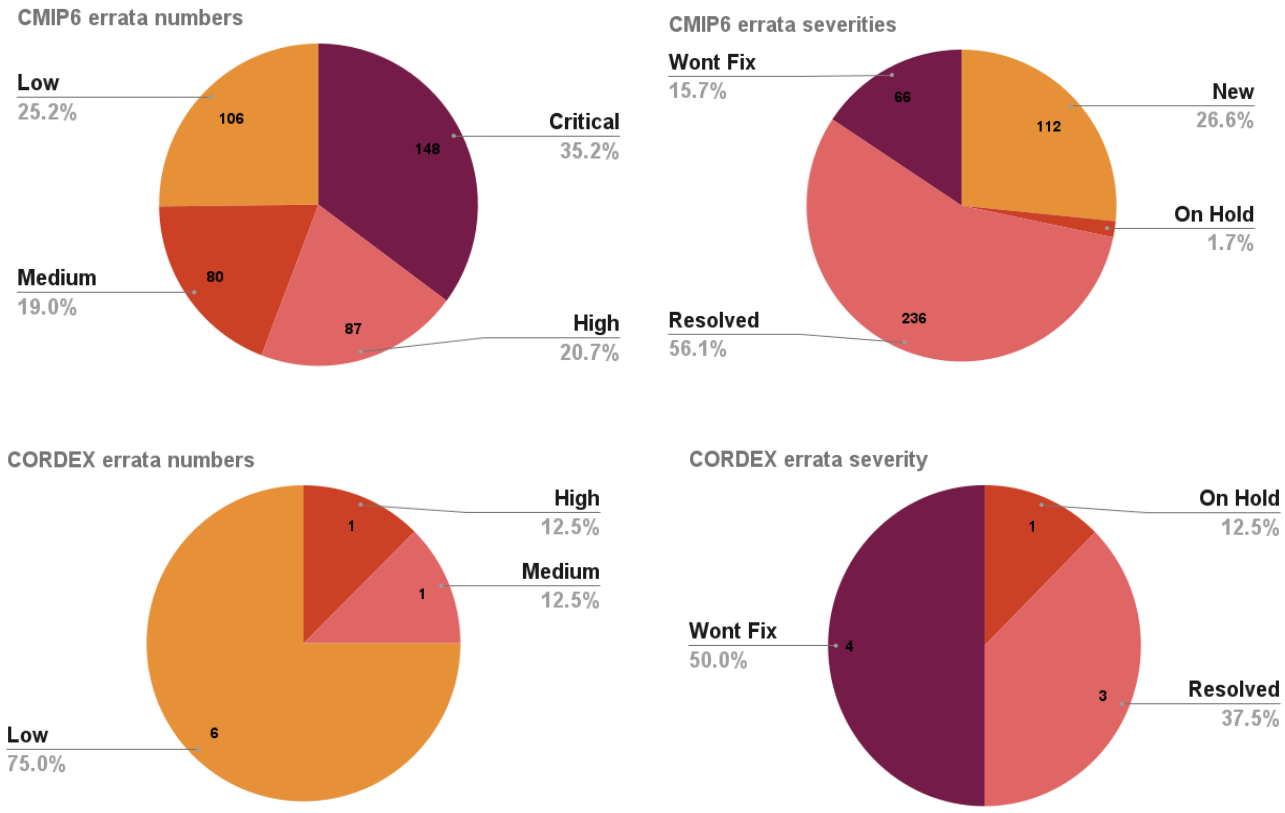


Figure 16. CMIP6 and CORDEX errata provisioning (numbers and severity),"Wont Fix" indicates issues which will not be fixed by the modeling groups

The ES-DOC model documentation service is designed for the modelling centres that need to provide their model documentation which is useful for a large range of users of model data. The service has two main objectives:

i) supporting the creators of model documentation and
ii) supporting the users of the documentation.

The creators are supported via the ES-DOC liaison at each CMIP6 modeling institute. This person (or persons) has been trained by ES-DOC and organizes the creation of documentation locally. Their primary contact with ES-DOC is via the support email (*support@es-doc.org*) and a dedicated ES-DOC-liaison mailing list.

CMIP6 documentation is provided through the ES-DOC website and accessible via the further_info_URL service, and currently comprises:

- **Models**: 41 models from 16 institutes (4 new institutes this reporting period)
- **Machines**: 16 institutes (10 new this reporting period)
- **Performance**: 5 institutes (5 new this reporting period)
- **Experiments**: 324 experiments are described (10 new this reporting period)
- **Conformance**: 5 institutes (5 new this reporting period)
- **Simulations**: The cdf2cim service that automatically collects CMIP6 simulation descriptions from CMIP6 datasets published to ESGF now has over 2 million simulation records, which are ready to be transformed into on-line documents, accessible via the further_info_URL service

Since January 2022 there have been 26 new queries on the ES-DOC support email, with a total of 81 exchanges. Of these issues, 60% were from users of the service, and 40% from content providers from within the CMIP6 and CORDEX communities. Nearly all of the issues have been solved, apart from one outstanding query relating to model documentation being created but not showing up on the ES-DOC website. This is a serious consideration that is still being looked into in April 2023.

5 Conclusions and next steps

The usage of the core data distribution services stayed at a high level. There is a continued high demand for CMIP6 data especially from non-EU users (EU users continue to rely on the exploitation of the CMIP6 data pools established as part of IS-ENES3). The demand for CORDEX data increased strongly. A complete new KPI collection and reporting system for the Climate4Impact (C4I) platform was developed for the new portal version officially released in 2023. Thus C4I KPI reporting concentrated on these new and refined KPI collection (thus covering only a part of the latest reporting period).

The IS-ENES3 data distribution services will continue beyond IS-ENES3 and also the close collaboration in the worldwide ESGF federation will continue (especially to sustain coordinated global ESGF service operations). IS-ENES3 partner institutions have agreed to provide in-kind contributions to enable this and further coordination will take place in the ENES-RI legal entity currently being established (as part of the IS-ENES3 sustainability activity).

Support for groups which were involved in the IS-ENES3 compute VA and TA activities will be continued at institutional level. The effort to provide compute services in IS-ENES3 (providing compute platforms as well as web processing services) is currently continued in national as well as European projects which want to enable data near processing capabilities for the wider Earth science community.

Also the activities related to improving the FAIRness of data hosted in the ENES data infrastructure (and the overall ESGF federation) will be continued based on institutional funding: thus the operations of the PID service as well as the DOI service is ensured beyond IS-ENES3. Especially the ongoing activity (not planned for IS-ENES3) to make DOI-PID relations added to the PID metadata visible via the OpenAire Scholix Hub, enabling the cross-linking of DOI and PID level identifiers will improve the FAIRness of the overall CMIP6 data collection.