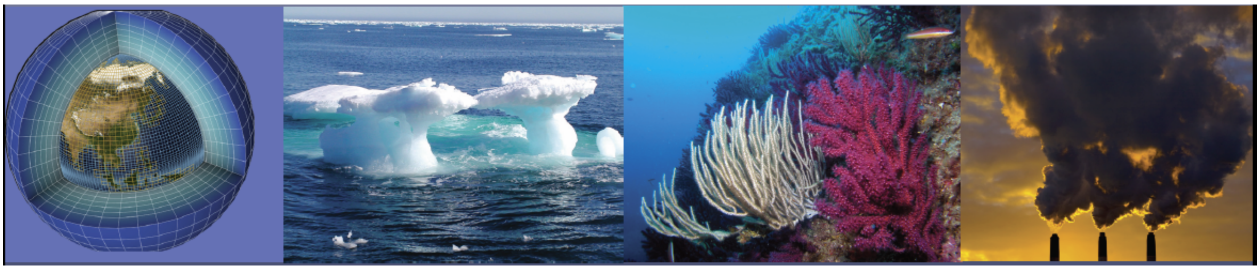
IS-ENES2

Project final report

Figures and tables



Partner n°	Partner **organisation name**	**Short name**	**Country**
1	Centre National de la Recherche Scientifique Institut Pierre Simon Laplace	**CNRS - IPSL**	France
2	Deutsches Klimarechenzentrum GmbH	**DKRZ**	Germany
3	Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique	**CERFACS**	France
4	Centro Euro-Mediterraneo per i Cambiamenti Climatici	**CMCC**	Italy
5	University of Reading National Centre for Atmospheric Science	**UREAD - NCAS**	United Kingdom
6	Met Office	**MetO**	United Kingdom
7	Science and Technology Facilities Council	**STFC**	United Kingdom
8	Sveriges Meteorologiska och Hydrologiska Institut	**SMHI**	Sweden
9	Koninklijk Nederlands Meteorologisch Instituut	**KNMI**	Netherlands
10	Max Planck Institute for Meteorology	**MPG**	Germany
11	Climate System Analysis Group, University of Cape Town	**CSAG**	South Africa
12	University of Manchester	**UNIMAN**	United Kingdom
13	Institutul National de Hidrologie si Gospodarire a Apelor	**INHGA**	Romania
14	Wageningen Universiteit	**WU**	Netherlands
15	Linköpings Universitet	**LiU**	Sweden
16	Barcelona Supercomputing Centre	**BSC**	Spain
17	Universidad de Cantabria	**UC**	Spain
18	Deutsches Zentrum Für Luft- und Raumfahrt in der Helmholtz Gemeinschaft	**DLR**	Germany
19	Danish Meteorological Institute	**DMI**	Denmark
20	Institut Català de Ciències del Clima	**IC3**	Spain
21	Météo France Centre National de Recherches Météorologiques	**MF - CNRM**	France
22	Universitetet i Bergen	**UiB**	Norway
23	Norwegian Meteorological Institute	**met.no**	Norway

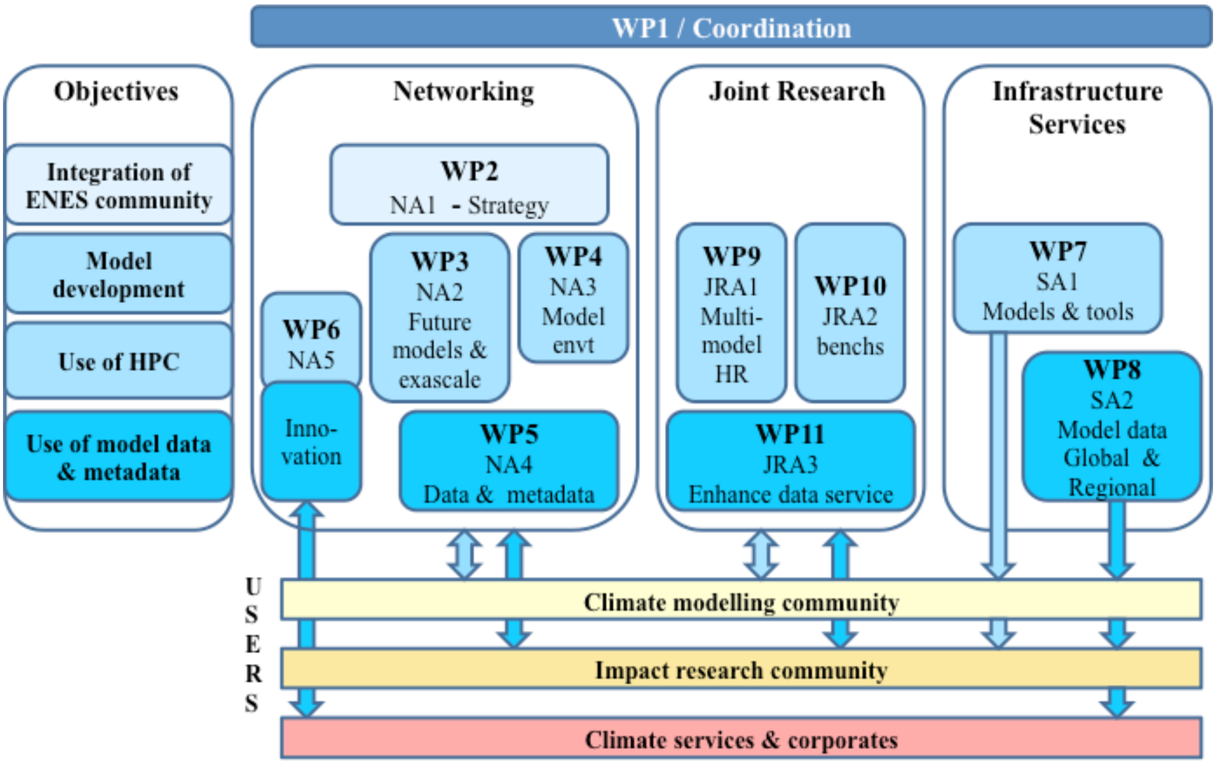


Figure 1: Work package activities with respect to the project's objectives and relations with user communities.

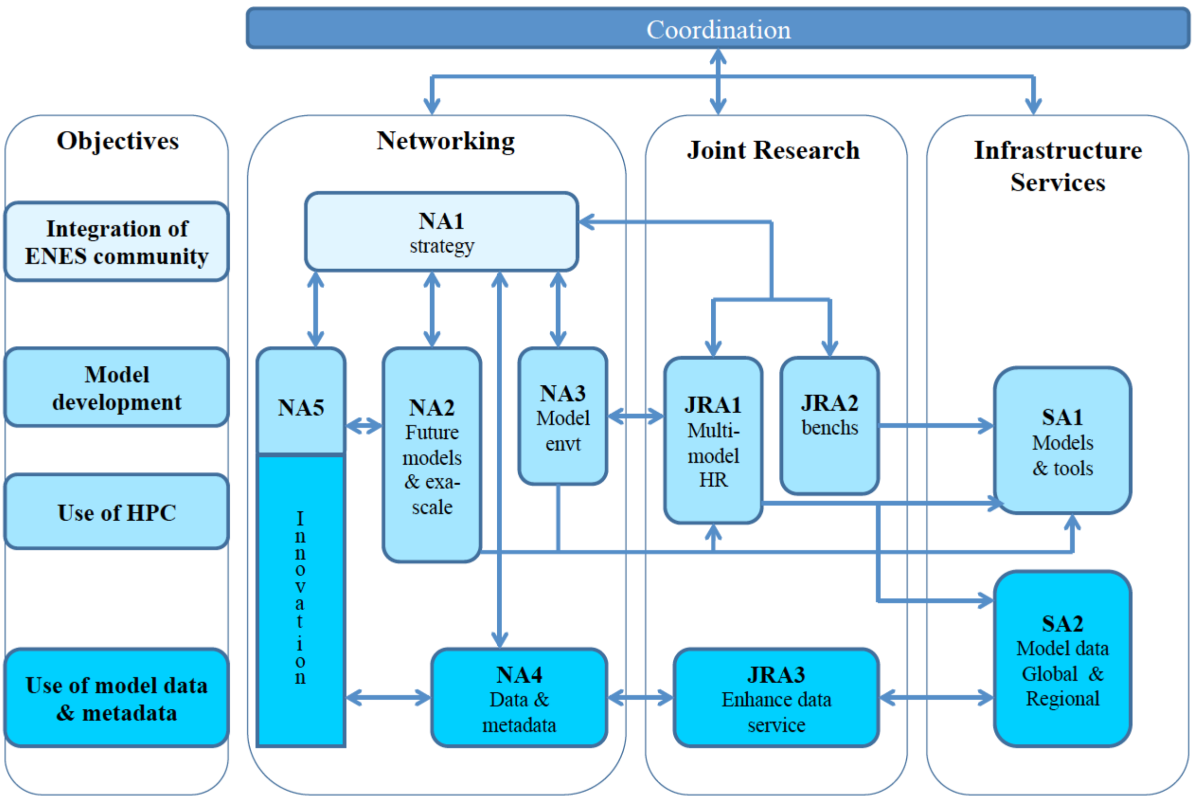


Figure 2: Relationship of work package activities to each other and the major networking, joint research, and infrastructure activities.

1. Foster the integration of the European climate and Earth system modelling community

	Recommendations
Models	Accelerate the preparation for exascale computing by exploiting next generation hardware *at scale* as early as possible, recognising that new algorithms, software infrastructures, and workflows will be necessary and will take substantial time and effort to develop.
High-performance computing	Work through national and European facilities to exploit a blend of high-performance computing facilities, recognizing the need to support both current and next generation science. Sustained access to world-class machines and next generation architectures will be needed to make a step-change in climate science.
Data	The community should invest more in research into data standards, workflow, data handling, high performance data management, and data analytics to meet the challenges of increasing data volumes and complexity. It should ensure that the systems, standards and workflows are in place, with sustainable funding and suitable mechanisms for establishing requirements, so to that data from climate simulations are easily available and well documented and quality assured, especially for downstream users.
Physical network	Work with national and international network providers to maximize the bandwidth between the major European climate data and compute facilities and ensure that documentation and guidance on tools and local network setup are provided for end-users and their local network administrators.
People	The community should be proactive about advertising the intellectual and technical challenges in climate science, both to individuals and to colleagues in other disciplines. Institutions should increase opportunities for training in climate science modelling and underlying technologies, at all levels from undergraduate to doctoral training courses and summer schools, as well as strengthen networking of software engineers.
Model evaluation	**NEW**: The community needs to put in place shared governance procedures for climate model evaluation software, which covers both the evaluation aims and the software structures. Data infrastructure should be expanded to include the computational resources needed for routine evaluation of new data products as they join the community ensembles.
Organisation	**NEW**: The European climate science community needs to utilise both national and European funding to develop sustainable community based cooperation that would put both production science and long-term future development activities on a firmer footing. In doing so, it needs to develop clear interfaces with other service providers (e.g. GEANT) and service consumers (e.g. the Copernicus climate services).

Table 1: Recommendations of the mid-term update of the ENES infrastructure strategy

2. Enhance the development of Earth System Models

	1rst Period (18 months)	2nd Period (18 months)[1]	3rd Period (12 months)
Version	OASIS3-MCT	OASIS3-MCT3.0	OASIS3-MCT3.0
Change sets	558	549	231
Downloads	230	>174	78
User support	416	>231	371
Tickets updates (Redmine)[2]	219	475	41

Table 2: Service activity on OASIS

	1rst Period (18 months)	2nd Period (18 months)[3]	3rd Period (12 months)
Version	CDO 1.5.9 to 1.6.4	CDO 1.6.5to 1.7.1	CDO 1.7.1 to 1.8.0
Downloads	> 2000	> 12434	9400
Support requests	> 2700	> 874	947
Tickets opened/closed[4]	-	72/64	76/84

Table 3: Service activity for CDO

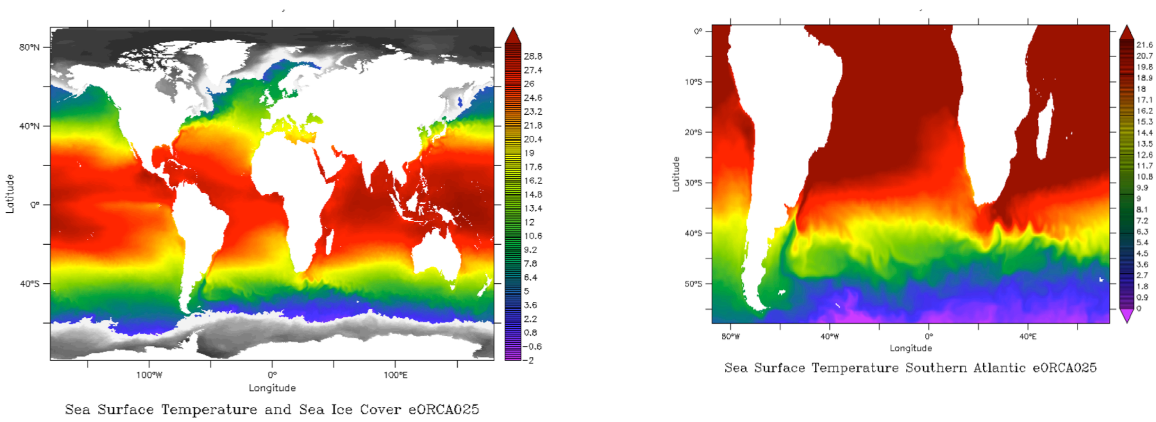


Figure 3: Sea surface temperature and sea ice cover from eORCA025 (1/4°) with a zoom on the Southern ocean (courtesy of CNRS-IPSL)

	1rst Period (18 months)[5]	2nd Period (18 months)[9]	3rd Period (12 months)
Changes of source code	> 744	> 1500	2000
Opened tickets/bugs fixed	233/140	322/326	65/45
Edits of the wiki pages	> 482	> 300	150
User support (mail)	> 300	> 140	350

Table 4: Service activity on NEMO

[1] Due to the computation of new, more meaningful KPI with semi-yearly frequency started during the second period (from month 9 onwards), some values relative to this period are given as a subset of the total as they only cover the last year of the 18 month period.

[2] https://inle.cerfacs.fr/projects/oasis3-mct

[3] Due to the computation of new, more meaningful KPI with semi-yearly frequency started during the second period (from month 9 onwards), all values relative to this period are given as a subset of the total as they only cover the last year of the 18 month period.

[4] This information is only available for part of the 2nd period and for the whole 3rd period.

[5] Due to the computation of new, more meaningful KPI with semi-yearly frequency started during the second period (from month 9 onwards), all values relative to this period are given as a subset of the total as they only cover the last year of the 18 month period.

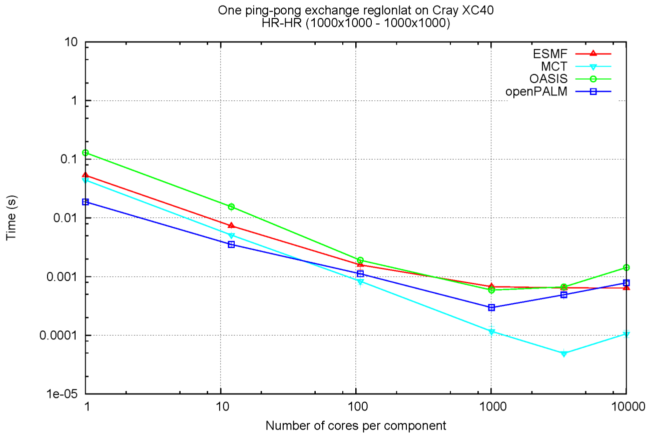


Figure 4: Benchmarking of four coupling technologies on the Cray XC40: time for one back-and-forth exchange between two components running on a regular latitude-longitude grid, representative of today high-resolution atmosphere and ocean models (from D10.3)

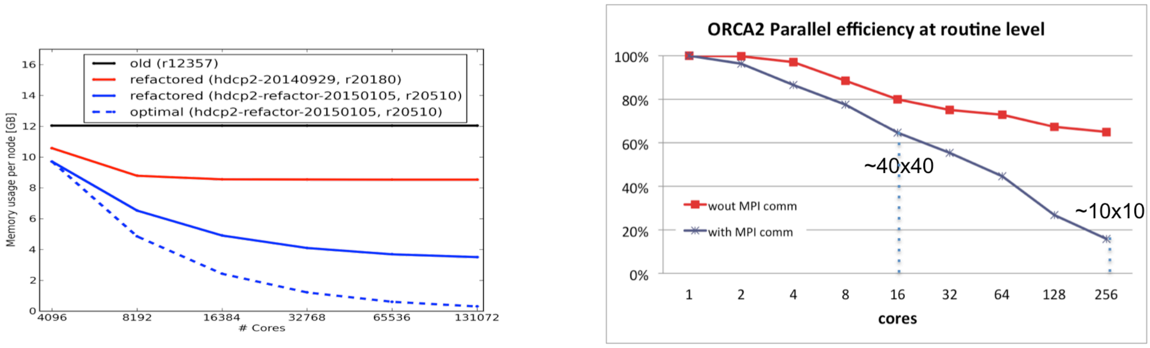


Figure 5: (Left): Optimization of the memory usage per compute node for a local area model with a resolution of 240m on the BlueGene/Q system JUQUEEN (courtesy of DKRZ); (Right): ORCA2 (200 km) parallel efficiency of routines performing (blue) or not (red) MPI communications[6].

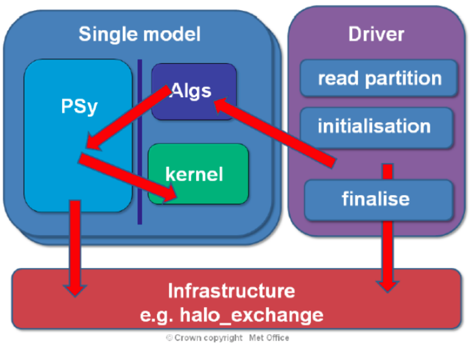


Figure 6: Separation of concerns between the Parallel Systems (the PSy layer) and Algorithms and Kernels[7]

[6] In the working document available on the project portal https://portal.enes.org/ISENES2/documents/na2-working-documents/the-nemo-oceanic-model-performance-analysis/view

[7] See Special report on IS-ENES2 web site "Application of the PsyKAl approach to the NEMO ocean model" Fort R. et al., 2017

Figure 7: typical workflow for model execution, post-processing and analyses. From MS4.7

3. Foster high-end simulations enabling to better understand and predict future climate change

Model and Platform	
Resolution	Number of grid points (latitude x longitude x vertical)
Complexity	Number of prognostic variables per component
Platform	Description of the computational hardware

Computational cost	
SYPD	Simulated years per day
ASYPD	Actual SYPD obtained from a long running simulation
CHSY	Core hours per simulated day
Parallelization	Total number of cores allocated for the run
JPSY	Energy cost of a simulation (Joules per simulated year)

Coupling/memory/I/O	
Coupling cost	Overhead caused by coupling
Memory bloat	Ratio of actual memory size to the ideal memory size (size of complete model)
Data output cost	Cost of performing I/O
Data intensity	Measure of data produced per compute hour

Table 5: Computational Performance Model Intercomparison Project (CPMIP) metrics (Balaji et al. (2017))

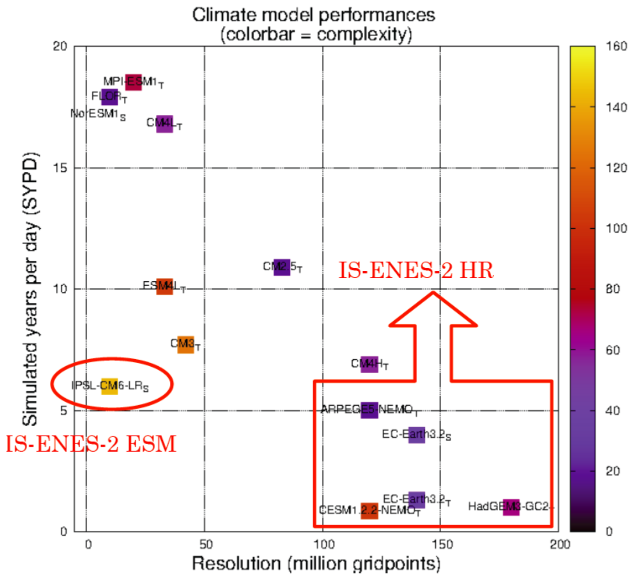


Figure 8: Computational performance in SYPD for European ESMs as a function of resolution (total number of grid points) and model complexity (color coded) (from D9.1)

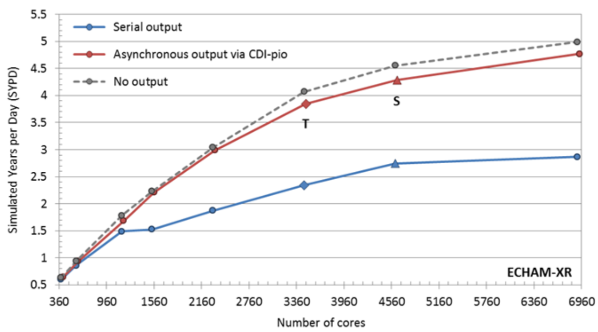


Figure 9: Simulated number of years per day (SYPD) for ECHAM-XR (T255L95) configuration with serial output (blue), asynchronous parallel output via CDI-pio (red) and without output (dashed grey) (From D9.4).

Models	First measurements (2014)	Final measurement (2016)
ARPEGE5-NEMO	13%	1%
EC-Earth3	24.7%	4%
HADGEM3-GC2	n/a	15%

Table 6: Coupling cost (% of total CPU cost) of HR models

Description	Type
MPI-ESM1	Coupled ESM used in CMIP5
IPSL-CM	Coupled ESM used in CMIP5
CMCC-CESM-NEMO	Coupled ESM in preparation to CMIP6
EC-EARTH	Coupled ESM used in CMIP5
ICON	Dynamical core
NEMO tracer advection	Kernel
ICON communication	Kernel
Coupling benchmark suite	Version 1.0.0

Table 7: List of benchmarks available on ENES portal

Vendors' perspective	**Consequences for strategy in ESM**
Heterogeneity of systems (mixture of CPU / GPU / accelerators) will increase	Increasing number of abstraction in memory levels have to be expected and will require serious code adaption to use the systems efficiently. This could include delegating system adaption to tools like OpenMP, DSL and/or subdivide models into software layers (abstraction) on top of optimized kernels (tuning level).
Increasing number of cores per socket with decreasing amount of local memory per core. Decreasing Memory Bandwidth and network bandwidth per core. Clock rate of cores cannot be expected to increase in the same rate as in the past. On the contrary, it will probably decrease.	Performance improvements can only be achieved by a higher level of parallelism, which then will increase code complexity, will substantially impede code development and efficient use of the new systems. Higher parallelism with at the same time minimal communication will be required and might narrow the choice of algorithms available depending on respective communication patterns.
Hardware development will increasingly focus the mass market (e.g. mobile phones) and emerging new market segments (e.g. deep learning) and decreasingly serve classical application areas with special requirements (e.g. climate science)	This may lead to reduced engagement in supporting special toolkits, compiler, libraries, higher support costs or to slower response times when it comes to bug fixing etc.

Table 8 summarizes the vendors' perspective on future architectures and their consequences for the strategy to be followed by ESMs.

4. Support the dissemination of Earth system model data to the climate and impact research communities

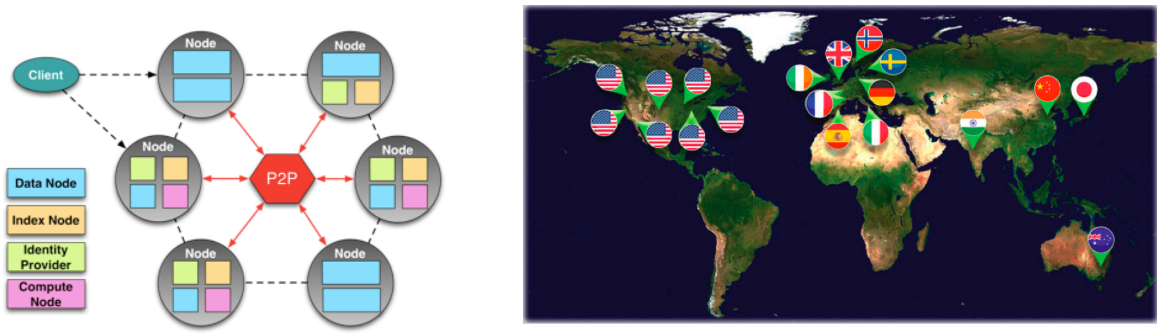


Figure10: (left) description of the peer-to-peer service of ESGF and (right) location of ESGF datanodes (6th ESGF F2F annual meeting report, 2017 and D11.5).

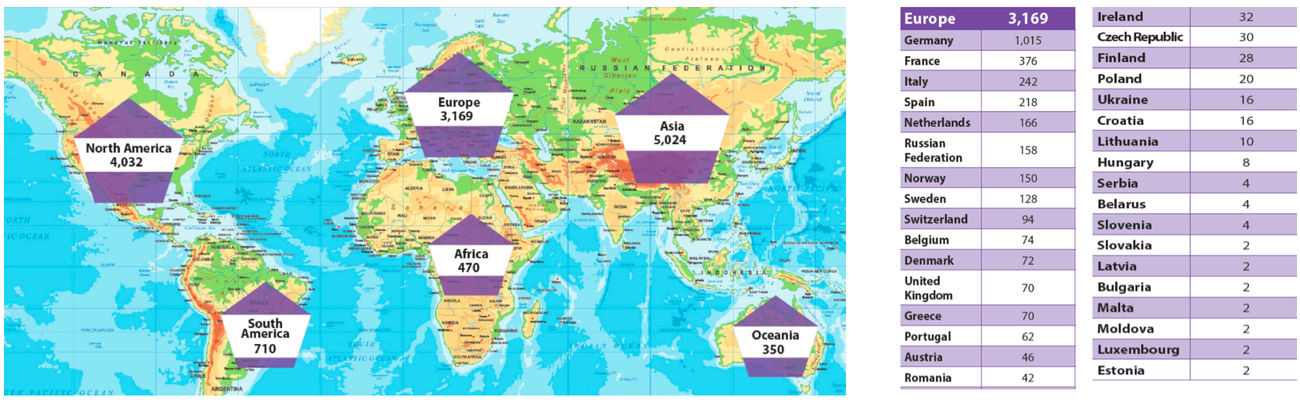


Figure 11: Distribution of ESGF registered users by continent and details for Europe (right) (6th ESGF F2F annual meeting report, 2017 and D11.5).

Host institution	ESGF datanode	Number of datasets	Institutes providers of data published
CEDA (UK)	esgf-data1.ceda.ac.uk	10108	MOHC (UK)
DKRZ (DE)	esgf1.dkrz.de	4717	MPI-M (DE)
IPSL (FR)	vesg.ipsl.upmc.fr	3933	IPSL (FR)
IDRIS (FR)	prodn.idris.fr	120	IPSL (FR)
LIU (SE)	esg-dn1.nsc.liu.se	1840	EC-EARTH group
ICHEC (IE)	esgf.ichec.ie	374	EC-EARTH group
Meteo France (FR)	esg.cnrm-game-meteo.fr	1075	CNRM-CERFACS (FR)
NorStore (NO)	noresg.norstore.no	591	NOR-ESM group (NO)
PCMDI (USA)	aims3.llnl.gov	280	CMCC (IT)

Table 9: List of CMIP5 datanodes hosting data from the seven European global climate models (from ESGF on 21/05/2017)

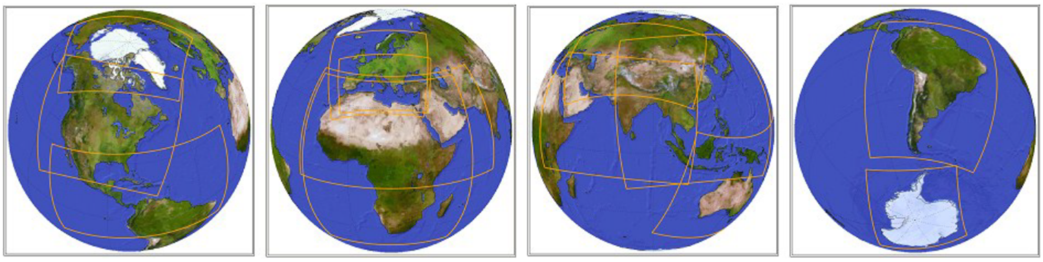
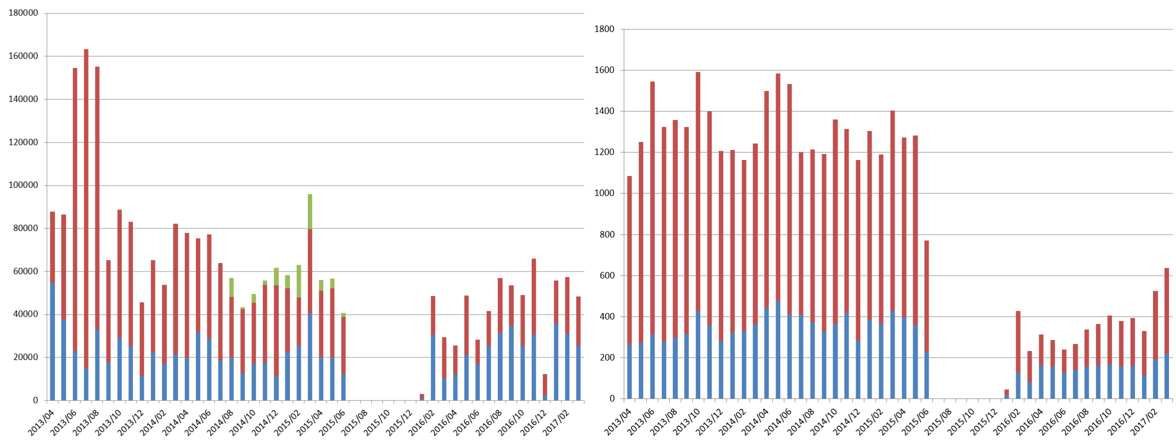


Figure 12: CORDEX domains

Institution hosting the datanode	**ESGF node**	**Number of datasets**	**Institutes providers of data published**
NSC-LIU (SE)	esg-dn1.nsc.liu.se	48543	SMHI (SE), ICTP (IT), MGO (RU) & CORDEX-Adjust
DKRZ (DE)	esgf1.dkrz.de	13441	CLM community (DE, CH), AWI (DE), DHMZ (HR), MPI-GERICS (DE), RMIB-UGent (BE)
DMI (DK)	cordexesg.dmi.dk	12928	DMI (DK), HMS (HU), KNMI (NL), ULg (BE), UQAM (CA)
CEDA (UK)	esgf-data1.ceda.ac.uk	2611	MOHC (UK)
NCI (AU)	esgf.nci.org.au	2550	UNSW (AU)
CCCR (IN)	esg-cccr.tropmet.res.in	2526	IITM (IN)
IPSL (FR)	vesg.ipsl.upmc.fr	929	IPSL-INERIS (FR) & CORDEX-Adjust
Meteo France (FR)	esg.cnrm-game-meteo.fr	422	CNRM (FR)
U. Cantabria (ES)	*data.meteo.unican.es*	*2500 (tbc)*	*UCAN (ES), CORDEX-Adjust,*
CSAG (S. Africa)			*CORDEX-ESD*

Table 10: Status of ESGF CORDEX datanodes (21/05/2017). In black, groups having run EuroCORDEX and some other domains, in blue institutes running other domains. In bold, IS-ENES data nodes (from ESGF CORDEX, 21/05/2017). Italic, datanodes for CORDEX-ESD ready but still to be opened when data are available.



Figures 13: Key performance indicators for ENES data infrastructure from 04/2013 to 03/2017: (Left) Volume of data downloaded (in GB) from ESGF European datanodes, (Right) number of active users on the nodes. Blue color for European users and Red non-European users, Green for use of WDCC archive.

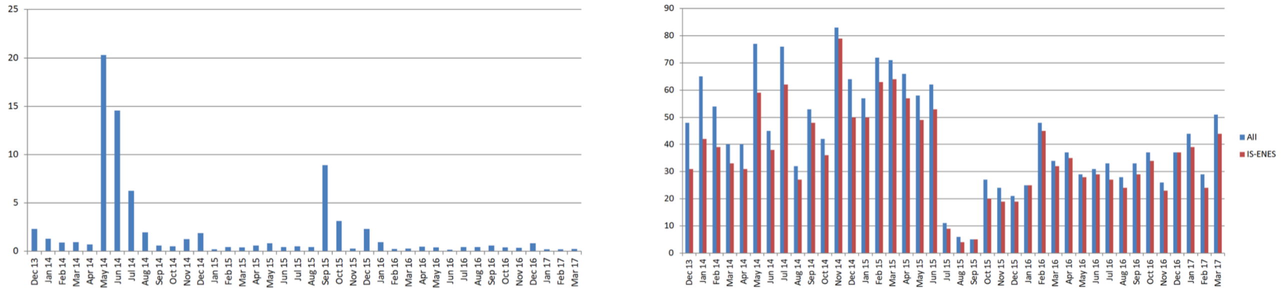
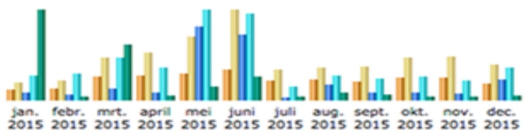


Figure 14: ESGF helpdesk from 12/2013 to 03/2017. (Left) Mean response time of ESGF user support (in days); (Right) Total number of email threads (blue) and IS-ENES2 contribution (red) (KPI).

Period	Unique visitors	Number of page hits
February 2016-December 2017	2290	8000
Jan 2017-March 2017	404	1595

Table 11: Statistics on the usage of ES-DOC model documentation



month	Unique visitors	Number of visitors	pages	hits	bytes
jan. 2015	1002	1644	43149	142263	48.08 GB
febr. 2015	1059	1855	32931	153179	1.90 GB
mrt. 2015	2186	3923	64718	247721	29.51 GB
april 2015	2255	4430	41533	187479	2.46 GB
mei 2015	2490	5920	429910	525534	7.25 GB
juni 2015	2843	8358	383313	505838	12.60 GB
juli 2015	1806	2819	15752	81211	2.01 GB
aug. 2015	1907	3027	91892	145005	4.20 GB
sept. 2015	1688	3099	42509	123899	2.82 GB
okt. 2015	2074	3924	43705	139625	2.08 GB
nov. 2015	2049	4019	39164	115474	1.85 GB
dec. 2015	1497	3317	122402	191622	2.50 GB
Totaal	22856	46335	1350978	2558850	117.26 GI

month	Unique visitors	Number of visitors	pages	hits	bytes
jan. 2016	1556	3063	193052	319625	4.21 GB
febr. 2016	1547	3113	106549	231020	3.55 GB
mrt. 2016	1652	3044	490162	763530	20.60 GB
april 2016	1633	2959	299685	453191	22.00 GB
mei 2016	1462	2670	192229	334389	6.88 GB
juni 2016	1488	2594	192661	368976	4.84 GB
juli 2016	1332	2427	4516235	4650347	43.40 GB
aug. 2016	1509	2848	373426	576147	33.43 GB
sept. 2016	1592	2852	257662	481896	71.68 GB
okt. 2016	1861	3395	258896	417680	198.58 GB
nov. 2016	1262	2230	176557	252651	14.09 GB
dec. 2016	0	0	0	0	0
Totaal	16894	31195	7057114	8849452	423.26 GB

Figure 15: Statistics of usage of the climate4impact portal for 2015 and 2016 (KPI).

Working Team	Working Team Leads	Team Goals
1. CoG User Interface Working Team	Cecelia DeLuca (NOAA) and Luca Cinquini (NOAA)	Improve ESGF search and data cart management and interface
2. Metadata and Search Working Team	Luca Cinquini (NASA)	Implement ESGF search engine based on Solr5 and discoverable search metadata
3. Publication Working Team	Sasha Ames (DOE) and Rachana Ananthakrishnan	Enable capability for publishing CMIP and other project data sets to ESGF
4. Node Manager Working Team	Sasha Ames (DOE) and Prashanth Dwarakanath (IS-ENES)	Manage ESGF nodes and node communications
4a. Tracking/Feedback Notification Working Team	Sasha Ames (DOE) and Prashanth Dwarakanath (IS-ENES)	Implement user and node notification of changed data in the ESGF ecosystem
5. Identity Entitlement Access Management Working Team	Philip Kershaw (IS-ENES) and Rachana Ananthakrishnan (DOE)	Implement ESGF X.509 certificate-based authentication and improved interface
6. Compute Working Team	Charles Doutriaux (DOE) and Daniel Duffy (NASA)	Develop data analytics capability within ESGF
7. Quality Control Working Team	Martina Stockhause (IS-ENES) and Katharina Berger (IS-ENES)	Integrate external information into the ESGF portal
8. Installation Working Team	Nicolas Carenton and Prashanth Dwarakanath (IS-ENES)	Install the components of the ESGF software stack
9. Dashboard Working Team	Paola Nassisi (CMCC) and Sandro Fiore (IS-ENES)	Monitor the Earth System Grid Federation in terms of system metrics and data usage statistics
10. International Climate Network Working Group	Eli Dart (DOE/ESnet) and Mary Hester (DOE/ESnet)	Increase data transfer rates between the ESGF climate data centers
11. Data Transfer Working Team	Lukasz Lacinski (DOE) and Rachana Ananthakrishnan (DOE)	Enhance ESGF data transfer and web-based download
12. Software Security Working Team	George Rumney (NASA) and Dan Duffy (NASA)	Implement security measures to identify vulnerabilities in the ESGF software and provide continuous improvement to the ESGF software development life cycle.
13. Support Working Team	Torsten Rathmann (IS-ENES) and Matthew Harris (DOE)	Develop frequently asked questions regarding ESGF and housed data
14. Documentation Working Team	Matthew Harris (DOE) and Sam Fries (DOE)	Document the use of the ESGF software stack
15. Replication and Versioning Working Team	Stephan Kindermann (IS-ENES) and Tobias Weigel (IS-ENES)	Create replication tool for moving data from one ESGF center to another; in addition, preserve versioning history of the ESGF published data sets
16. Provenance Capture Working	Bibi Raju (DOE)	Enable ESGF provenance capture for

Table 12: ESGF Working teams leadership, from ESGF 1^{s}t Implementation Plan (2016)

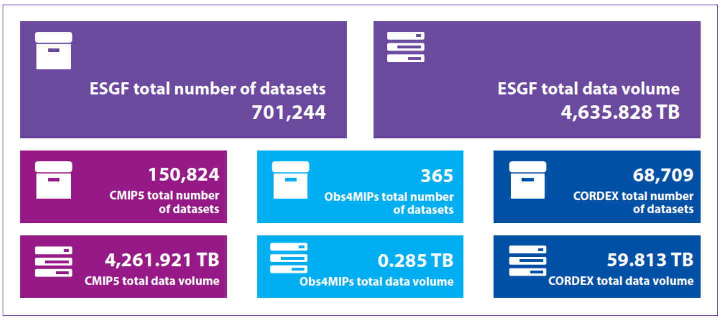


Figure 16: ESGF data archive statistics obtained from the ESGF Console (from the 6th ESGF F2F annual meeting report, 2017 and D11.5). Contributions from CMIP5, Obs4MIP and CORDEX are detailed. IS-ENES2 supports the European contribution to CMIP5 and CORDEX

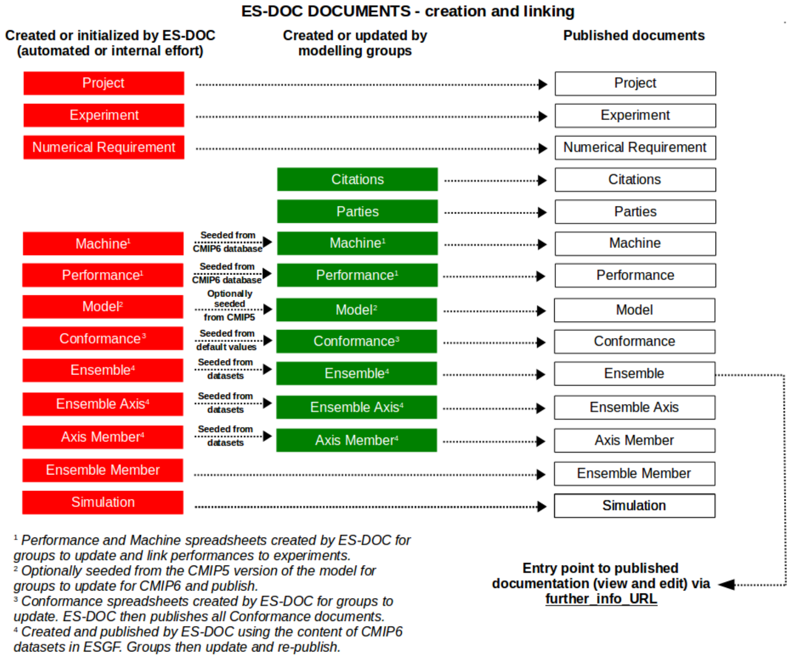


Figure 17: Overview of ES-DOC documents and creation workflow for CMIP6. The right hand side lists all documents and their relationships. The documents listed on the left are generated by ES-DOC (either automatically or via internal effort). The central column lists the documents either created by modelling groups (Citations, Parties, Machine and Model) or updated toward their final version. Latest version and complete CMIP6 documentation ecosystem can be found on http://es-doc.org/cmip6.

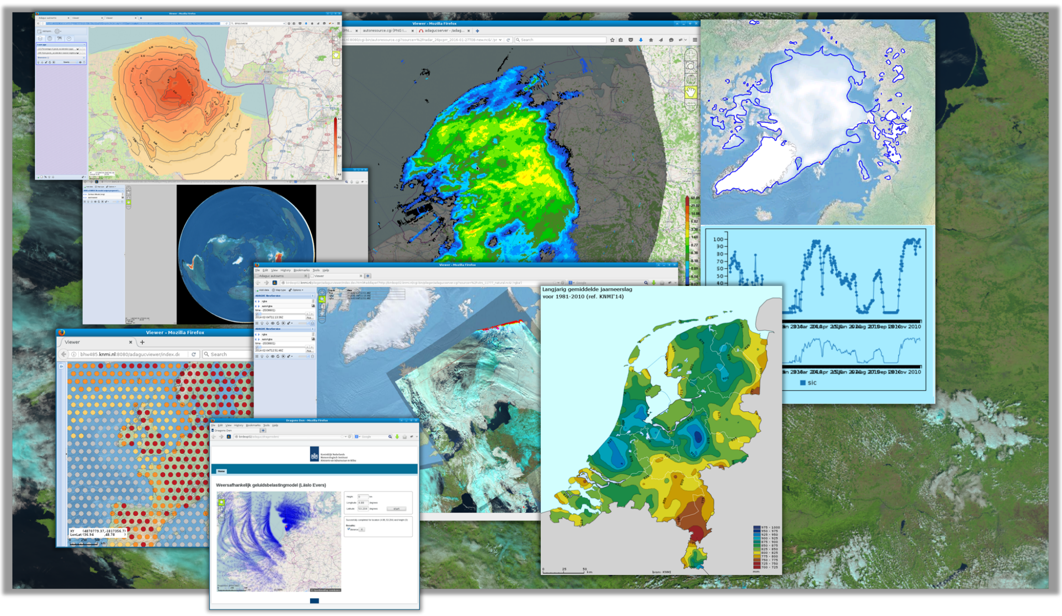


Figure 18: Different functionalities offered by the climate4impact portal (from Page et al., talk at the 6th ESGF F2F meeting, December 2016)