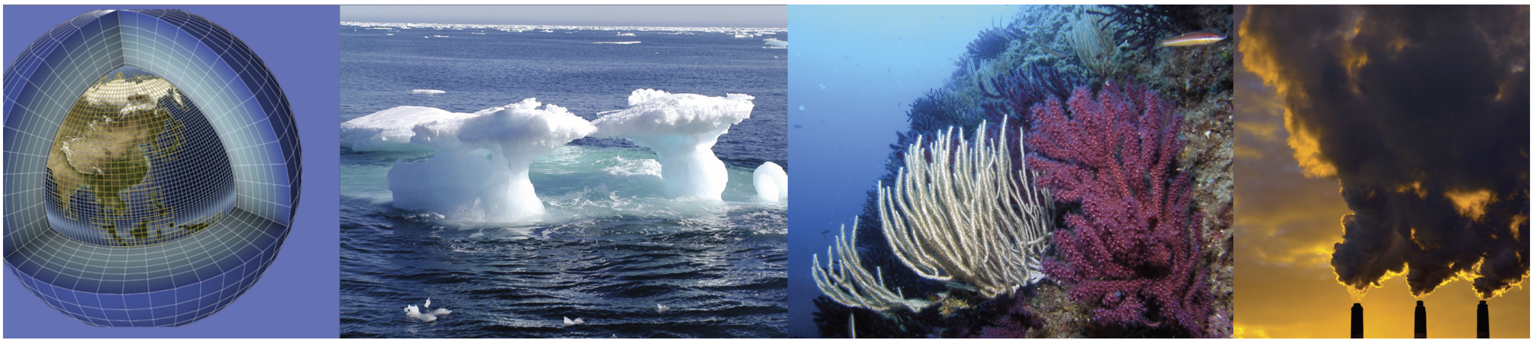
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

enes
EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

Update of the
ENES infrastructure strategy
2012-2022



2017

Update of the ENES infrastructure strategy 2012-2022

2017

Authors:
Sylvie Joussaume (IPSL, FR), Bryan Lawrence (NCAS, UK), Francesca Guglielmo (IPSL, FR)

Contributors:

Mario Acosta (BSC, ES)
Giovanni Aloisio (CMCC, IT)
Jean Claude André (consultant, FR)
Joachim Biercamp (DKRZ, DE)
Olivier Boucher (IPSL, FR)
Björn Brötz (DLR, DE)
Anca Brookshaw (ECMWF, UK)
Reinhard Budich (MPIMet, DE)
Mick Carter (MetOffice, UK)
Antonio Cofiño (U. Cantabria, ES)
Sébastien Denvil (IPSL, FR)
Francisco Doblas-Reyes (BSC,ES)
Ralf Döscher (SMHI, SE)
Bernd Eggen (MetOffice, UK)
Veronika Eyring (DLR, DE)
Kerstin Fieg (DKRZ, DE)
Sandro Fiore (CMCC, IT)
Rudolf Fischer (NEC, DE)
Rupert Ford (STFC, UK)
Marie Alice Foujols (IPSL, FR)
Eric Guilyardi (IPSL, FR)
Christoph Heinze (U. Bergen, NO)
Bart van den Hurk (KNMI, NL)
Colin Jones (MetOffice, UK)
Martin Juckes (BADC, UK)
Christopher Kadow (FUB, DE)
Michael Kolax (SMHI, SE),
Julian Kunkel (DKRZ, DE)
Michael Lautenschlager (DKRZ, DE)
Claire Lévy (IPSL, FR)
Grenville Lister (NCAS, UK)
Eric Maisonnave (CERFACS, FR)
Jochem Marotzke (MPIMet, DE)
Alastair McKinstry (ICHEC, IE)
Silvia Mocavero (CMCC, IT)
Steve Mullerworth (MetOffice, UK)
Christian Pagé (CERFACS, FR)
Philipp Neumann (DKRZ, DE)
Tim Palmer (U. Oxford, UK)
Marie Parinet (IPSL, FR)
Michael Rezny (Monach U., AU)
Kim Serradell (BSC, ES)
Øyvind Seland (MetNo, NO)
Alistair Sellar (MetOffice, UK)
Bjorn Stevens (MPIMet, DE)
Rowan Sutton (NCAS, UK)
Claas Teichmann (GERICS, DE)
Frank Toussaint (DKRZ, DE)
Sophie Valcke (CERFACS, FR)
Manuel Vega (U. Cantabria, ES)
Xavier Vigouroux (ATOS, FR)

Citation reference:
Joussaume S., Lawrence B. and Guglielmo F. (2017), Update of the ENES infrastructure strategy 2012-2022, *ENES Report Series 2*, 20 pp

Table of contents

Executive Summary

The ENES infrastructure strategy developed during the first phase of the IS-ENES project (*Mitchell et. al. 2012*) outlined a drive towards convective scale global modelling with improved initialisation and larger ensemble sizes as a grand challenge, providing clear infrastructure requirements. It specifically recognized that the infrastructure needed to deliver on that programme would also be that needed to deliver on advances in other aspects of climate science, including, but not limited to, the attribution of climate change, enhancing paleoclimate modelling, and the consideration of climate predictability at regional scales. That strategy provided key recommendations for the ENES infrastructure.

The second phase of IS-ENES, the IS-ENES2 project, started implementing the ENES infrastructure recommendations: setting goals, determining relevant actions, and screening for available resources. In doing so it committed to an interim review of progress and requirements. For this purpose, discussions were held within IS-ENES2 in a dedicated workshop (October 2016) and during the Final General Assembly (January 2017), leading to this document, the mid-term update of the 2012-2022 ENES infrastructure strategy.

The 2012 strategy is updated with an analysis of drivers and infrastructure components in the context of 2017. This effort confirms previous recommendations on models, HPC, data, networks, and people with partial reformulations based on a combination of progress and evolving requirements. It complements them with guidance on scientific evaluation of models and emphasizes the need to organizationally tackle the sustainability challenge, reflecting the approach of a more mature community.

The 2017 recommendations are:

1) **On models**: Support common development and sharing of software and accelerate the preparation for exascale computing by exploiting next generation hardware and developing appropriate algorithms, software infrastructures, and workflows.

2) **On HPC**: Exploit a blend of national and European high-performance facilities to support current and next generation science and work toward obtaining sustained access to world-class resources and next generation architectures.

3) **On model data:** Evolve towards a sustained data infrastructure providing data that are easily available, well-documented and quality assured, and further invest in research into data standards, workflow, high performance data management and analytics.

4) **On physical network:** Work 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 available to users.

5) **On people:** Grow the numbers of skilled scientists and software engineers in the ENES community, increase opportunities for training at all levels, and strengthen networking between software engineers**.**

6) **On model evaluation (new):** Enhance sharing of common open source diagnostics and model evaluation tools, implement governance procedures, and expand data infrastructure to include computational resources needed for more systematic evaluation of model output.

7) **On infrastructure sustainability (new):** Sustain the cooperation necessary to develop future model and data technology and support international reference experiments programmes, and strengthen collaboration with other European actors providing services to, or using services from, ENES.

1. Introduction and objectives

In 2012, the European Network for Earth System Modelling (ENES) published an "*Infrastructure Strategy for the European Earth System Modelling Community*" based on meetings held in 2010 and 2011 (*Mitchell et al., 2012*).

This strategy addressed the underlying needs for the delivery of the next decade of European research working on climate change and variability. It outlined a drive towards convective scale global modelling, with improved initialisation and larger ensemble sizes – recognising that the infrastructure needed to deliver on that programme would also directly support work on the attribution of climate change, on enhancing paleoclimate modelling, and on climate predictability at regional scales.

The key recommendations for the decade to come for the ENES infrastructure were to:

1. Provide a blend of high-performance computing facilities ranging from national machines to a world-class computing facility suitable for climate applications, which, given the workload anticipated, may well have to be dedicated to climate simulations.
2. Accelerate the preparation for exascale computing, e.g. by establishing closer links to PRACE[1] and by developing new algorithms for massively parallel many-core computing.
3. Ensure data from climate simulations are easily available and well documented, especially for the climate impacts community.
4. Build a physical network connecting national archives with transfer capacities exceeding Tbits/sec.
5. Strengthen the European expertise in climate science and computing to enable the long-term vision to be realized.

Since 2012, there has been considerable progress on these objectives, much with the support of the InfraStructure for the European Network for Earth System Modelling project (IS-ENES2)[2]– funded by the European Commission under FP7. Activities that have been taken forward range from the European engagement and leadership in the Earth System Grid Federation (ESGF[3]) providing access to climate model data, to the collaboration on high-performance computing (HPC) that paved the way to the EC funded European Centre of Excellence in the Simulation of Weather and Climate ESIWACE[4], a first step in preparing Earth system simulation for next generation computing. However, many of the existing activities are being sustained by relatively short term funding (e.g. ESIWACE), even though they are meant to be long-term activities; wider recognition of the need for the long-term sustainability of European research infrastructure for climate modelling is still needed.

Five years on, it is important to ask whether the 2012 infrastructure strategy still represents the main scope of the ENES community as it did five years ago. Are these still the right objectives, is the community making progress towards them, and what would long-term sustainability of the infrastructure look like?

This document was prepared within IS-ENES2 project. It represents the output of a meeting held in October 2016 where representatives of the ENES community gathered with the express aim of addressing these questions and providing a formal mid-term update to the 2012-2022 strategy. The recommendations were further discussed at the project final General Assembly in 2017.

In this mid-term update we do not repeat all the arguments laid out in *Mitchell et al. (2012)*, rather we revisit the requirements of the ENES infrastructure strategy, beginning with an updated view of the

---

[1] PRACE: Partnership for advanced computing in Europe (http://www.prace-ri.eu)

[2] IS-ENES2: FP7 infrastructure project (2013-2017), following IS-ENES FP7 project (2009-2013) (http://is.enes.org)

[3] ESGF Earth System Grid Federation (http://esgf.llnl.gov/)

[4] ESIWACE: H2020 Centre of Excellence project (2015-2019) "Excellence in Simulation for Weather and Climate" (https://www.esiwace.eu/)

societal and scientific drivers. We then proceed illustrating the status of the infrastructure as of 2017. The different components of the infrastructure are analysed, progress and requirements are discussed, identifying changes in landscape and activities with respect to the previous strategy. Finally, we propose an updated roadmap, with a new set of recommendations that, while building on the previous ones, reflect the approach of a more mature community.

We anticipate that this updated strategy will be of use at three levels:

1. To indicate to those responsible for making European funding decisions how European investment in infrastructure to support Earth system simulation and related data provision could deliver rewards to European society that build on, but do not replace, national investments.
2. To help colleagues making national funding decisions understand how their decisions impact not only on their national priorities, but also on the synergies possible at the European level.
3. To help colleagues across our discipline (including those in relevant national and European institutions) understand:
    - The relationship between scientific goals and the necessary infrastructure,
    - Our collective inter-dependencies on infrastructure both now and in the near (up to five years) future, and
    - The relationship between the costs and risks of joint approaches and the potential added value.

2. Societal and scientific drivers in 2017

The ENES community initially encompassed primarily those who developed and used models of the Earth's climate system to understand climate variability and change under natural and anthropogenic forcing. However, with time the scope has broadened. It is now inclusive of all of climate modelling science, from regional to global and seasonal, decadal and centennial scales; for prediction, projection, and process understanding; covering Earth system physics, chemistry and biogeochemistry. With increasing interest in "seamless" approaches to modelling across scales, and convergence of tools, ENES has necessarily also engaged in strengthening ties with the numerical weather prediction community. This increase in scope has arisen from both scientific necessity, and societal requirements, both of which, as emphasized in 2012, drive the community and its infrastructure.

2.1 Societal drivers

In 2012 it was already clear that the development of strategies of mitigation and adaptation to climate change themes needed to address themes such as vulnerability and risk across society. Addressing these themes required, and still requires: (i) on-going targeted efforts towards model improvement and efficient data provision and (ii) ever better interfaces with the wider climate services community. Both require technical and scientific advances to be entrained into an infrastructure available to a wider community than the initial practitioners.

Since 2012, these requirements have only grown. The adoption of the Paris Agreement in 2015 reinforces the need for climate science to provide knowledge addressing mitigation and adaptation policies.

Mitigation of climate change requires better insight into the biogeochemical processes and the carbon cycle for greenhouse gas emission verification and possible negative emissions from land-use change. Adaptation requires more understanding of risk at relevant local scales – leading to the need for improvements in generating and managing model ensembles (which address understanding uncertainty), in the resolution and fidelity of models at regional scales, and in better downscaling methods.

Both adaptation and mitigation require the provision of projection data and information to a wide range of users. The need for climate information tailored to different sectors (food, water, energy, ecosystems, health, economy) has significantly increased and climate services are developing at both national and European scales. Multi-model reference climate simulations are recognised as part of the whole chain going from climate knowledge to the delivery of tailored climate information to users (*Street et al., 2015*). They rely on the internationally coordinated experiments organised by the World Climate Research Program under the Coupled Model Intercomparison Project, complemented by the Coordinated Regional Climate Downscaling Experiments (CORDEX).

In Europe, Copernicus has launched a new Climate Change Service (C3S) – ENES was intimately involved in precursor activities and is actively involved in supporting the C3S. The ENES initiative precursor to C3S was the CLIPC[5] (Climate information platform for Copernicus, 2013-2016) project that offered services in terms of access to climate change and variability information (observed and projected), and transformed data products enabling impact assessments and development and manipulation of impact indicators for Europe. CLIPC served a wide variety of users from scientists, to the public and private sector. CLIPC illustrated the possible contributions from the ENES infrastructure to C3S.

Several ENES modelling groups cooperate with C3S, participating in prototype development in different directions: facilitating access to and manipulation of global and regional projections using IS-ENES expertise, generating products, fidelity metrics and tools, extending global/regional model combinations and scenarios to reduce uncertainties, and working towards a reference set of climate projections for Europe from the perspective of sectoral applications (preparing the operational phase of C3S).

This progressive growth in importance of climate services bears an impact on a number of the conclusions as well as on the requirements and the road towards sustainability of infrastructure as delineated in 2012.

2.2 Scientific drivers

The scientific drivers have not changed substantially since 2012. They remain the necessity to address climate predictability on a range of timescales, understand climate sensitivity and major feedbacks, improve reliability of simulations at regional scale, understand past glacial-interglacial cycles, and better understand how climate is changing due to anthropogenic forcing.

Progress with respect to these questions has been achieved in recent years, with contributions from European funded ENES projects. COMBINE[6] and SPECS[7] have contributed to a better understanding of climate predictability. EUCLIPSE[8] has contributed to better understanding of the role of low-level clouds in climate sensitivity uncertainty. EMBRACE[9] has advanced understanding of Earth system processes. On-going EU H2020 projects include PRIMAVERA[10] addressing the role of spatial resolution to improve regional reliability of climate simulations and CRESCENDO[11] focusing on

---

[5] CLIPC: FP7 Copernicus pilot project (2014-2016) on “Climate Information Portal for Copernicus” (https://www.clipc.eu)

[6] COMBINE: FP7 Project (2009-2013) “Comprehensive Modelling of the Earth System for Better Climate Prediction and Projection” (http://www.combine-project.eu/)

[7] SPECS: FP7 project (2012-2016) “Seasonal-to-decadal climate Prediction for the improvement of European Climate Services » (http://www.specs-fp7.eu/)

[8] EUCLIPSE: FP7 project (2010-2013) “European Union Cloud Intercomparison, Process Study and Evaluation project” (http://www.euclipse.eu/)

[9] EMBRACE: FP7 project (2011-2015) on “Earth system model bias reduction and assessing abrupt climate change” (http://www.embrace-project.eu/)

[10] PRIMAVERA: H2020 project (2015-2019) on “Process-based climate simulations: advances in high-resolution modelling and European climate risk assessment” (http://www.primavera-h2020.eu)

[11] CRESCENDO: H2020 project (2015-2019) on “Coordinated research on Earth system and climate: experiments, knowledge, dissemination and outreach” (http://www.crescendoproject.eu)

improving the representation of terrestrial and marine biogeochemical as well as natural aerosol and trace gas processes.

Major scientific issues emphasized in 2012 in the ENES infrastructure strategy:

- How predictable is climate on a range of timescales and what are the limiting factors? Can the range of uncertainty be fully represented with the models we have available, and without exaggerating the range of possible futures?
- What is the sensitivity of climate and how much can we reduce the current uncertainty in the major feedbacks, including those due to clouds, atmospheric chemistry and the carbon cycle?
- What is needed to provide reliable predictions of regional changes in weather and climate?
- Can we model and understand glacial – interglacial cycles, including changes in carbon cycle and major ice sheets? Can we use observational evidence from past climates to calibrate the sensitivity of complex climate models and respective adjustable model parameters?
- To what extent can we attribute signals in the period of the instrumental record to understand Earth system processes – from weather scales to those typical of anthropogenic climate change?

While significant progress has been achieved, these issues are as yet unresolved, and new ones have arisen. For example, the need to establish frontier challenges around the importance of where the carbon goes, how weather changes with climate and how climate influences habitability, arose from the 2015 Paris negotiations (*Marotzke et al., 2017*).

These issues are fully consistent with the launching of seven grand challenges by the World Climate Research Programme (WCRP) (*Brasseur and Carlson, 2015*) covering: (1) clouds and atmospheric circulation and how they relate to climate sensitivity, (2) how carbon feedbacks affect the climate system, (3) understanding and predicting weather and climate extremes, (4) improving near-term climate prediction, (5) how water availability will affect food, (6) how the cryosphere will respond to and affect climate change, and (7) how sea-level will rise and affect coastal regions. These challenges reflect areas where the international scientific community and its stakeholders consider that major progress needs to be made. They give more emphasis to climate extremes and the cryosphere compared to ENES 2012 scientific questions, as a result of five years of progress in climate science.

These grand challenges constitute a main component of WCRP strategy to accelerate progress in climate science – requiring the community to exploit their full spectrum of expertise in the coming 5 to 10 years.

2.3 Consequences

The climate modelling community has to simultaneously deal with:

- Running current models and facilitating the use of the associated data – to progress understanding of climate, improve models, and inform society;
- Preparing for future generation models and data exploitation running on future computer architectures.

These two streams naturally result in activities with different workflows and timescales. Running and improving current models, sharing and delivering data and producing meaningful aggregated products to both climate science and climate services, and progressing near term climate science depend on reliable infrastructure, short turn around in model development, and global collaboration. Preparing for the future requires a very different approach: focused technical development over very long-timescales exploiting co-design of both software and hardware aimed at producing systems and codes which are ready for next generation challenges. These two streams and their requirements are described below.

State of the art Climate Science: The WCRP CMIP cycle

The internationally coordinated CMIP experiments form a set of reference simulations, which are important for model evaluation, understanding of processes, and provision of scenarios for future climate change. These simulations are extensively used by the climate research community, but also by other communities. Indeed, complemented by the downscaled experiments CORDEX based on CMIP experiments, they serve as a reference to the communities studying the impacts of climate change, and more and more for climate services.

CMIP is now in its sixth phase, CMIP6**.** The WCRP grand challenges constitute the backdrop of the CMIP6 (2015-2020), with a focus on three broad questions (*Meehl et al., 2014, Eyring et al., 2016a*):

1. How does the Earth system respond to forcing?
2. What are the origins and consequences of systematic model biases?
3. How can we assess future climate changes given climate variability, climate predictability, and uncertainties in scenarios?

Europe plans to contribute to CMIP6 as described in section 3.

CMIP experiments have major infrastructural requirements: up-to-date reference versions of models need to be developed, and the many experiments require large allocations of dedicated computing time on stable architectures as well as the archiving of large amounts of data. Moreover, in order to ease coordinated analyses, data and metadata standards have to be defined and imposed, and an international infrastructure for online data storage and distribution with common rules and policies is required (the Earth System Grid Federation, ESGF). The two phases of EU funded IS-ENES have demonstrated the added value of working together at the European scale to support CMIP and CORDEX and develop and deploy the ESGF to manage and distribute data products.

Next Generation Climate Modelling

The 2012 infrastructure strategy emphasized a grand challenge for climate modelling. An explicit representation of deep cloud convection in atmospheric models would make a major step in the representation of processes. This would require increasing the global spatial resolution of such models to 1 km, revisiting aspects of the physics, and in the accompanying ocean models necessitate resolving ocean eddies. This vision for “1-km modelling” translates into a major dimensioning challenge for the infrastructure needed to deliver highly scalable climate models ready for future extreme computing architectures.

The establishment of a new centre of excellence for weather and climate computing (ESIWACE) in late 2015 was a community response to address the first steps towards solving the technical challenges required for 1km global (weather or climate) modelling.

In 2016, the European climate community reaffirmed this grand challenge by proposing a large programme on extreme climate and computing aimed at improving the resilience of Europe to climate change[12]. The proposed programme made apparent the link between developing 1km scale global climate models, the requisite extreme computing, and the provision of quantitative estimates of the changing character of climate extremes. The requisite computing could only be provided by future exascale computing platforms, which, being very different in nature from existing platforms, would require completely new modelling methods. Such methods would need a long-term joint European

---

[12] EPECC H2020 Future and Emerging Technology flagship proposal (https://ec.europa.eu/futurium/en/content/flagship-european-programme-extreme-computing-and-climate), which has evolved into the ExtremeEarth flagship (htpps:// http://www.extremeearth.eu/)

approach including both the weather and climate modelling communities, and build upon the work of ESIWACE. Whether or not this particular programme goes ahead, the community is committed to take such steps.

As in 2012, in proposing these activities, the community recognises that much of the infrastructure needed to deliver 1km scale models will also be that needed to deliver on other necessary advances in climate science such as enhanced complexity and additional model components (e.g. ice sheets).

3. Status of the Infrastructure in 2017

Climate modelling depends on a significant amount of infrastructure: from the models themselves to the computing and data infrastructure. High performance computing is necessary to run simulations, and large-scale archive and data analysis systems are needed to exploit and distribute the data. More and more emphasis is also given to the overall workflow to run simulations and access model results, the physical network required to exchange data, all the diagnostic software required to analyse model results and evaluate their quality. This tripod of models, computing, and data – with all the connecting workflow, network, and tools – needs to be well documented and depends on the central role of people with expertise in climate science, model and software development (Figure 1).

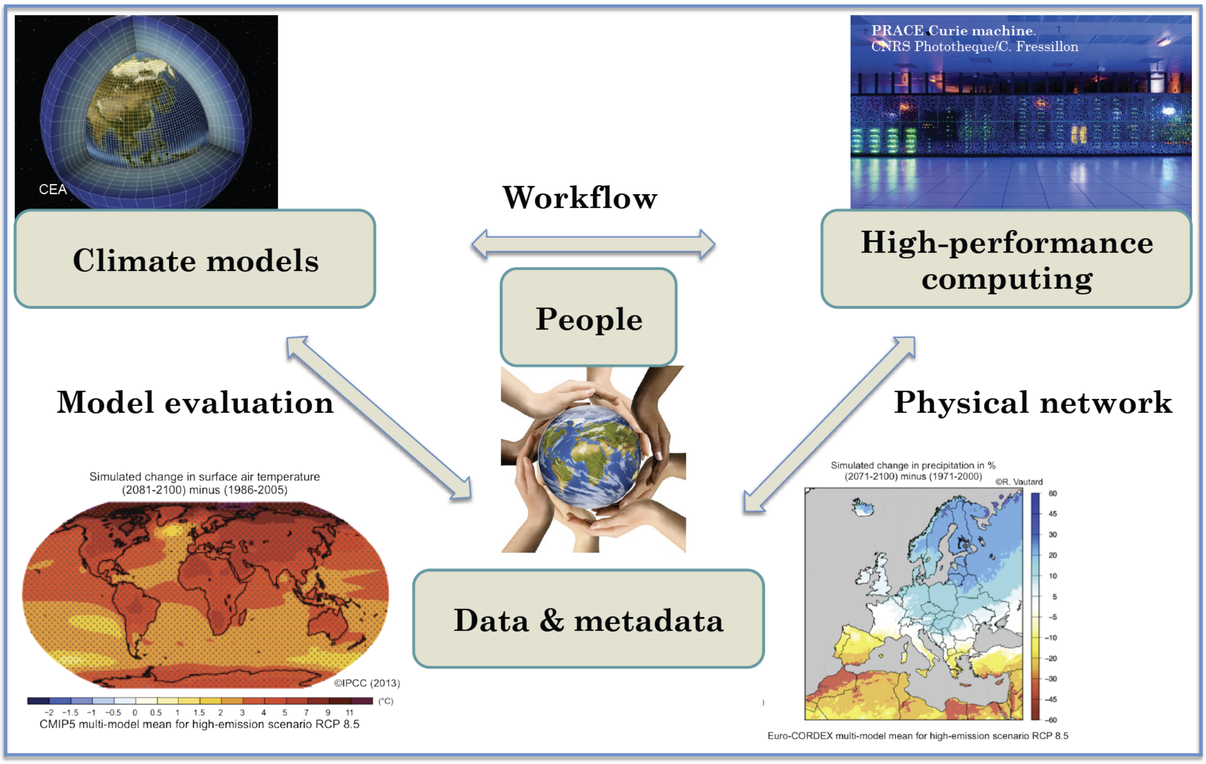


***Figure 1**: the main components of climate modelling infrastructures*

In the following, we take each of these components in turn, and describe their current status and evolution since 2012.

3.1 Models

The overall European landscape in climate modelling remains similar to the one described in 2012. Seven European model families have contributed to the CMIP5 and will continue to be the major European contributors to CMIP6: from the EC-Earth consortium, France (2 model families), Germany, Italy, Norway, and the United Kingdom. Since CMIP5, the scope has broadened and two additional German models will join CMIP6, focusing on atmospheric chemistry and ocean modelling, respectively.

The seven major European model families share some components, such as the European ocean modelling platform (NEMO) used within five of the models. Many also use the same coupler (OASIS) and/or the same IO server (XIOS). The overall landscape brings some diversity, which, as was emphasized in 2012, is needed to provide some understanding of the influence of model representation on uncertainty. However, whether this represents the optimal balance between efficiency (sharing of the technical burden) and scientific diversity is still under debate. It is unknown whether the existing multiple different European models provide significantly more information about model structural uncertainty than would be obtained from different combinations of model complexity and resolution, as done for example with the different ocean configurations used by the NEMO consortium. Nevertheless, it is now clearer that differences in much of the underpinning infrastructure provide no scientific benefit, and it would be desirable to have a cleaner separation of concerns between scientific diversity and the underlying software infrastructure where much more could be in common. However, it is also recognised that some diversity in the underlying software is necessary to provide evolutionary pressure so that the software itself can evolve to become more efficient and capable of exploiting more computational environments.

The 2012 roadmap recommended: *to strengthen European collaboration for model development, prepare for future computer architectures and improve model parameterizations.*

The main development of codes and parameterisations is carried out at the national level, within modelling groups and centres. However, ENES favours common development of new parameterisations through joint science projects, and of shared software and tools through joint infrastructure projects (such as the IS-ENESx family). Within IS-ENES2, the most emphasis was put on improving the performance of tools such as coupler and IO servers, on promoting the development and use of common components, on promoting standard indices for intercomparing model performance, and on sharing of best practices. Also within IS-ENES2, for the first time the level of shared understanding as to how to parameterise radiation was such that a joint code could also be worked upon alongside radiation diagnostics.

Preparing for future computer architectures requires revisiting model codes to be able to run efficiently on a very large number of massively parallel, possibly hybrid, processors, alongside work on other elements of the infrastructure. There are two elements involved in making progress with model development: demonstrating the need, and preparing the codes. In terms of need, ESIWACE is developing an exascale exemplar, which will be further explored in the context of upcoming programmes. Regarding codes, important efforts have been made at national level to develop and apply new dynamical cores, which are more efficient to run on massively parallel architectures. However, these efforts are not heavily resourced and may not produce the required changes in time to readily exploit next generation computers, in part because most groups are constrained in effort; they need to continue to develop and improve existing code bases against short-term goals, and cannot afford to spend all their effort on next generation codes. A way forward is for more community sharing of software, with community support and common coding standards as exemplified by the NEMO consortium.

3.2 High-performance computers

High-performance computing is integral to climate modelling, and climate models make use of all the facets of HPC from massively parallel computations using high-bandwidth, low latency interconnect, to the use of high-performance storage systems. The choice of model resolution and complexity as well as the design of model experiments in terms of lengths and number of experiments is always a compromise with existing available computing power. To improve on any or all of spatial resolution, representation of complexity, number of experiments, ensemble size, duration of experiments, or assimilation, calls for increases in computing power. In practice, climate models need both capacity facilities, allowing for a large number of smaller experiments at limited resolution, and capability facilities, allowing the running of fewer large jobs such as small ensembles at high resolution.

The 2012 roadmap recommended *to develop access to world-class computers, develop the interface with the European HPC ecosystem and start collaborate with industry.* Progress has clearly been made in these directions.

National HPC facilities continue to be upgraded and heavily used. At the high end, the community has been proactively interacting with PRACE over the last 5 years, emphasizing needs for multi-year access and data storage. PRACE has supported opportunities to run extreme cases such as some of the first global atmospheric simulations at 25 km, very large ensembles of ocean at 25 km, predictability studies, and very long simulations in the Holocene.

The vision of the HPC ecosystem with its 3 tiers: Tier 0 world-class facilities, Tier 1 at national level and Tier 2 for institutional domain-specific facilities, as described in 2012, is evolving. Differences between those three levels are now seen less in terms of architecture or size, but more in terms of how large, and for how long, resources are allocated, and how data handling is supported, as well as how tiers are governed. The necessity for long-term continued access to stable platforms, with appropriate queues and data systems, is fundamental to climate modelling. Currently PRACE does not meet all these requirements and usage remains limited to few extreme simulations – a more intensive production mode would require one or more dedicated facilities.

Future architectures are expected to be more complex, with more hybrid and disruptive technologies and a multiplicity of architectural configurations. Working with industry in co-design mode will most probably be required to ensure systems adapted to applications. ESiWACE has started to tackle this issue with, for the first time, a European collaboration with relevant industry. Ensuring that codes perform when ported between architectures will become harder, and new coding techniques will be necessary. First experiences of domain specific languages to provide "separation of concerns" (separating the science code from the infrastructure code) show that it may be necessary to deeply modify the structure of codes (*Schulthess, 2015*).

3.3 Model data

Data infrastructure is an increasing concern in the field of climate modelling – both for the direct storage of data as it is produced in HPC facilities, and for the exploitation of the data by the scientific and wider communities. Volumes of data are increasing rapidly, with major European modelling centres expecting to be storing in excess of an exabyte of data each, early in the next decade (well before they will be dealing with exascale computing), and handling up to petabytes (PB) of data in analysis workflows which currently peak in the terabytes (TB). CMIP5 required the production of around 20 PB of data in Europe, contributing 500 TB to the globally shared ESGF CMIP5 repository (total of 2 PB). CMIP6 is expected to require the production of an order of magnitude more data (the UK alone expects to contribute 3 PB of CMIP6 data to ESGF). Alongside storage requirements, the volume of data and number of simulations are also both leading to new requirements for documentation standards and systems.

The 2012 roadmap recommended actions *to integrate distributed databases and contribute to international standards, develop interoperability with observations, and develop interface with impact community.*

The ENES community has taken additional leadership roles in supporting the development and operations of the Earth System Grid Federation (ESGF), and in doing so enabled the successful, on-going provision of CMIP5 data to the community. Under IS-ENES, the ENES community now drives half of the ESGF working teams, contributing to leadership at both the working and governance level. IS-ENES also provides most of the leadership and effort for the ES-DOC metadata systems aimed at documenting the CMIP experiments, the models used, and the simulations produced. Additional IS-ENES effort underpins support for the Climate-Forecast conventions for netCDF and for the CMIP6 Data Request. All these contributions produce added value by exploiting the collaborative opportunities associated with using European funding to leverage national funding in an international context – generating significantly more influence than the individual contributions alone.

Beside climate model results, the ESGF hosts reanalysis and satellite observation data, both to aid dissemination, and to support the direct comparison of simulations with observations using common formats. But still ESGF does not hold all data needed by climate sciences. Moreover, ESGF is too complex to use for many communities – the data volumes and tools can be intimidating, and many derived products (such as climate indices) are not available and cannot be derived without downloading the database. There is a need to develop climate analysis platforms for both climate and interdisciplinary research, allowing crossing data of different origin and lowering barriers of data accessibility. To address this issue for the community studying impacts of climate change, IS-ENES has developed the climate4impact portal[13] as a platform to explore data, access documentation and guidance, and perform predefined computations. This platform has been expanded within CLIPC to enable and facilitate access to model results (hosted on ESGF) for a wider community. Additional climate services have also been delivered in CLIPC, and work is underway on delivering ESGF data into the Copernicus Climate Change service. The participation of ENES in the EUDAT[14] consortium also aims at easing access to climate data to a much larger community. The European context of the European Open Science Cloud[15] may also bring some further opportunities.

The ESGF software needs to be further developed to meet increased needs in terms of performance, scalability, robustness, and easiness of operation, potentially exploiting new data handling technologies emerging from the commercial world. Sustained data delivery systems require long-term reliable funding and common, transparent applications rules and policies, and while climate service funding may contribute, future model intercomparison projects (CMIP6 and beyond) require more sustained pan-national support on top of existing national funding[16]. Such funding needs to recognise the increasing relative costs of data handling (relative to compute costs) associated with increasing data volumes. Workflows for both production and experimental simulations need to be improved. In particular, for the case of international model intercomparisons, the workflow needs to reliably proceed from model experiment definition, passing through data requirements, specifications, and data delivery, to data distribution systems such as ESGF, and climate services applications. Throughout, better support for underpinning standards would allow far more effective use of the data by a wider community. Further research into both data handling and data analytic technologies is needed.

3.4 Physical networks infrastructure

Climate science and climate services both depend on the aggregation of data from multiple sources such as model simulations carried out at disparate locations and observations produced by different

---

[13] htpps://climate4impact.eu

[14] EUDAT "Research data services, expertise and technology solutions" gathers IT experts on data as well as field applications (http://eudat.eu)

[15] European Open Science Cloud initiative (https://ec.europa.eu/digital-single-market/en/european-open-science-cloud)

[16] The importance of long-term sustained funding is also emphasized in the US Academy of Science's 2012 strategy for advanced climate modeling: https://doi.org/10.17226/13430

space agencies. The volumes of data and the number of data objects are large, and need to be moved in a timely manner – in some cases during or directly after the simulation. Expensive extra local copies of data can be avoided by using the distributed European archive as a distributed backup system. However, distributed data access (including distributed data backup) requires a reliable physical network, with low latency and high bandwidth, at affordable cost.

The 2012 roadmap recommended to *build a physical network connecting national archives with transfer capacities exceeding Tbits/sec* in order to ease transfer of data from computing centres and between the distributed climate model data nodes.

This requirement was based on extrapolating network bandwidth changes, but it now seems unlikely that Tbit/s will be available in the next few years. However, more data is now available as to requirements: Projections of data transfer requirements from existing projects (such as PRIMAVERA) suggest that 10 Gbit/s networks can cope if they only have to serve one sustained modelling project at a time – but this is unlikely over the next five years, particularly for data archives which take data from multiple HPC sites, or serve thousands of users. There we can expect the bandwidth requirements to reach up to 100 Gbit/s in the near future.

The headline bandwidth is only one part of the story. While key data sites are linked by high bandwidth national research/education networks and the international GEANT network, end to end bandwidth has not necessarily met the baseline provision. As a consequence, the ENES community has carried out network evaluation and testing in the context of two international activities: the International Climate Network Working Group, set up to measure and monitor data transfer performance between key sites, and the replication testing within the ESGF. Despite theoretical speeds of 10 Gbit/s, the largest sustained data transfers have been at around 5 Gbit/s, with more typical speeds being 1-2 Gbit/s. The typical speed can be even lower for extra European transfer (US West Coast, Australia, Asia). The main benefit of this work has been the recognition that the base backbone was not always the major limiting factor; the "last mile" is also integral to performance. The advent and deployment of "Science-DMZ"[17] areas on the edges of local networks has significantly improved some data transfer experiences and is something that should be more widely deployed.

Regardless of the base bandwidth into and out of large sites, it will be even more important to improve documentation for end-users and network administrators on how to exploit high-bandwidth data transfer technology. It will also be important for large sites to return to caching high-volume data, to avoid unnecessary re-transmitting data and -for sites with large caches of data - to support local computation to the greatest amount possible (as well as deploy local storage systems which can absorb/deliver high-volume inputs/outputs).

3.5 People

Climate modelling requires a blend of scientific and technical expertise, spanning climate and computer science as well as software engineering. However, much of the requisite expertise is also in demand across industry, supporting both the ubiquitous demand for data science skills and the requirements of technology companies such as Google, Amazon et al. – and supply is not meeting demand.

The 2012 roadmap recommended: to *strengthen the network of science experts and software engineers, enhance common developments, and develop training.*

This recommendation is still very relevant. Recognising that it is difficult to employ individuals with requisite skills, the main action within ENES has been to foster training and inter-institutional knowledge sharing. IS-ENES has explicitly supported both summer schools for young scientists and

---

[17] A science DMZ, https://fasterdata.es.net/science-dmz/, provides a network zone on the periphery of a campus or institution where network performance can be enhanced at the cost of some relaxation of security policies.

the networking of software engineers. The biennial summer schools were developed to introduce early career scientist and programmers to Earth System Modelling, to provide insights into how models work and their use to support science, and to make participants familiar with intercomparison exercises. Many workshops have provided venues for the sharing and exchange of practice for software engineers. Trans-national coding sprints providing shared development opportunities have been particularly efficient.

Climate science as a career needs to be made more attractive, recognising the need for a range of skills, from software engineering to computer science and mathematics, to climate modelling and climate science per se. However, the situation with respect to recruitment is unlikely to drastically improve, and so it will be important to find ways to entice new entrants to climate science, and to up skill those already in the field. Sustaining and extending training activities will be important In particular, training scientists to gain new skills to exploit new IT technologies and computing architectures will be essential, otherwise the inability to compete with industry in recruitment will limit the community ability to respond to a rapidly changing computational environment.

The community needs to better share experience within and between groups, ensuring technical transfer is prioritised as highly as scientific communications. The climate science community should intensify the dialogue with parallel disciplines to maximise transfer of information and solutions, with, as a possible outcome, the acknowledgement by computer science of the challenges and opportunities associated with the resolution of climate modelling problems and the exploitation of best practices from computer science by climate modellers.

3.6 Model evaluation common software

The use of climate models to extend scientific understanding of processes or to provide useful projections of the future is predicated on knowledge of model quality. Such foundational knowledge can only be constructed by confronting model simulations with observations – past and current - and with each other.

Since 2012, it has become apparent that the software and compute infrastructure required to make the necessary model-to-data and model-to-model comparisons requires significant intellectual effort to develop and maintain – especially in such a way that it can be readily used by third parties not involved in its development. A 2015 survey[18] of five European modelling centres developing atmospheric components clearly emphasized that model evaluation has a strong potential for more collaboration and for a reduction of duplication of effort. Initial developments in that direction have begun with the EMBRACE and CRESCENDO projects contributing to the development of the ESMValTool[19] for sharing well-established common model diagnostics.

The importance of such diagnostic and post-processing software, and access to the concomitant observational datasets, is now recognised, as is the expectation that evaluation of simulations should become more routine and feed back into model development – potentially exploiting a dedicated compute infrastructure alongside the ESGF (*Eyring et al., 2016b*). However, sharing model evaluation diagnostics and compute infrastructure will require appropriate new governance mechanisms, still to be developed. In particular a clear - preferably standard - interface between the science choices of model evaluation and their technical implementation needs to be established.

---

[18] JPI Climate, pan-European survey of the climate modelling community, Assimila, 2016 - http://www.jpi-climate.eu/media/default.aspx/emma/org/10875543/Survey_Final+report+on+European+ESM+V2.0.pdf

[19] (http://www.esmvaltool.org/)

4. Bringing these things together: ENES in the next decade

Infrastructure plays a crucial role to support ENES goals, both for running current models and for developing next generation models. It is also clear that infrastructure, by definition, requires long term investments to meet long-term strategic requirements - and this is best achieved by global collaboration across scientific communities.

For these reasons, the 2012 roadmap recommended to: *develop the long-term European infrastructure for climate modelling and to strengthen the role of Europe in international collaboration.*

The support to the infrastructure of ENES, delivered by two phases of the IS-ENES projects, has directly achieved many of these goals: strongly enhancing European contributions to the international data infrastructure for CMIP; expanding access to regional modelling results from CORDEX; and providing more direct support to other scientific communities such as those working on climate change impacts. IS-ENES support also developed, and has sustained, a network of software engineers sharing best practice, and has enabled wider use and understanding of European climate models by promulgating documentation and information via ENES portals. The IS-ENES projects have also provided the foundational inspiration for the advent of ESiWACE.

The successes of the IS-ENES projects made even more obvious the necessity to address more longer-term, sustained support for maintaining and evolving climate modelling infrastructure. As both the demand for climate model products and the complexity of the modelling challenge grow, relatively short term funding coupled to *ad hoc* collaboration around "*funding opportunities*" is no longer fit for purpose. Such funding methods risk the delivery of essential infrastructure, with the likely consequence that the expectations of society on the climate science community will not be met.

A first attempt to establish a sustainable European research infrastructure for the Earth's climate system modelling showed that definition and scope still raise a number of issues. However, these are not unique to Europe: at the international level, the concept of infrastructure is only now emerging with the advent of the CMIP6 infrastructure panel set by the Working Group on Coupled Models (WIP) alongside new international governance procedures for ESGF set in 2015.

This experience demonstrates the need for clear understanding both inside and outside the climate science community as to what can be achieved now, and what is necessary (in terms of both funding mechanisms and organisation) to meet expectations. With such agreement, it is possible that episodic funding could be supplemented by multilateral agreements that could arise from existing and future relationships between institutions and provide a mechanism for delivering the requisite long-term vision and sustainability. Such agreements would, however, need to recognize the need for distributed teams on episodic funding to regularly synchronize activity.

The eventual infrastructure will need to address the two major infrastructural challenges that need long-term sustenance: supporting production science using current models and developing next generation modelling systems. The former focuses primarily on standards, documentation, workflow, evaluation tools, data provisioning, data processing, and compute and storage systems. At the international level it is strongly driven by CMIP coordinated experiments that require agreed standards and a strong infrastructure organisation to run and analyse the results. The second challenge, addressing next-generation climate modelling is about the connection between science, software engineering, and vendor engagement. Together these two infrastructural challenges need to be overcome to address the WCRP grand challenges over the coming decade, with the aspiration of 1 km global modelling being an additional driver to providing the necessary scientific tools to take the community forward. Both require cross-community cooperation and both have data and HPC foci. Those two streams are complementary and both needed for climate science (Figure 2)

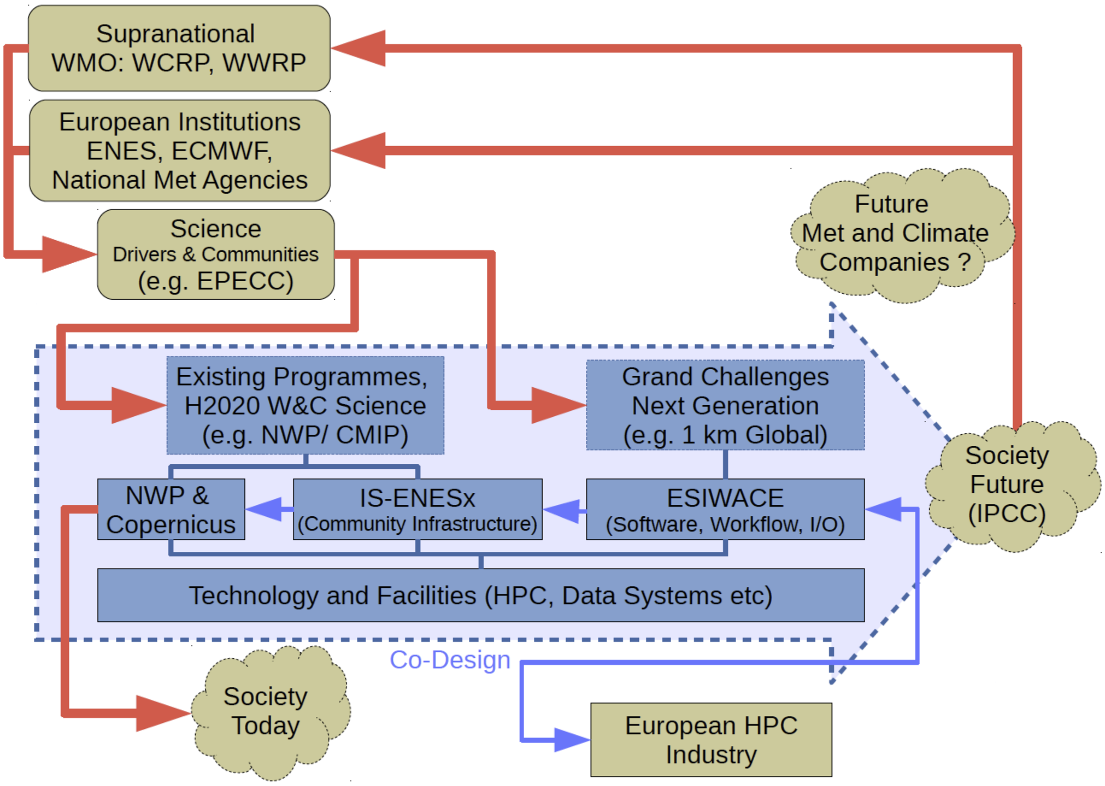


Figure 2*: Infrastructure for Earth's Climate System Modelling in Europe and beyond: main actors, drivers, projects, and initiatives intertwine at various scales and address the two main infrastructural challenges, sustaining CMIP production and preparing for extreme high resolution.*

Key to delivering the necessary infrastructure and collaboration are mechanisms to achieve community cohesion around aims and aspirations. Climate science needs to "*speak with one voice*" with respect to what can and cannot be delivered, and what needs to change to make delivery possible. The two ENES task forces, for HPC and data, are important instruments to elaborate and convey joint statements and advocate for common causes. Recent work of the ENES HPC task force has established fruitful interaction with PRACE and there are on-going efforts of the ENES data task force to coordinate ENES climate data infrastructure and to represent climate science in cross-disciplinary discussions on data infrastructures.

5. Recommendations - 2017

Based on the analysis and discussion carried out above, we present a revised and updated version of the recommendations.

5.1 Updated Recommendations in 2017

Recommendation (i): Models

ENES will continue to foster common development and sharing of software, such as model components, parameterizations or environment tools, through its science and infrastructure projects, e.g. IS-ENES. ESiWACE (and/or successor projects) need to push ahead with ensuring that climate and weather codes are always considered as crucial exemplars in European exascale projects, working with vendors in co-design where possible. Within ESiWACE and other projects, work on scalability and next generation codes needs to accelerate, recognising that a scalable dynamical core is just the first step, physics and data handling will also need to perform at exascale. However, funders will have to recognise that fast progress in this area is difficult and may involve false starts, and in doing so, reward an element of risk taking.

Revised Recommendation: Support common development and sharing of software. 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.

Recommendation (ii): High-performance computing

Climate science is still significantly limited both by capacity of available HPC, and by community access to capability high-end machines with suitable architectures. Providers of shared services need to better support the typical applications and workflow requirements of the community (including longer periods of access). Dedicated HPC facilities need better network bandwidth to dedicated data storage and analysis facilities.

The ENES community will need to continue to develop and maintain an HPC strategy via the work of the HPC task force. This strategy will need to recognise the importance of exploiting both national and trans-national resources to support current science, as well as delivering and shaping long-term access to the largest possible next generation machines to address grand challenge objectives.

Revised Recommendation: 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.

Recommendation (iii): Model Data

The development, maintenance and delivery of data systems for coordinated numerical experiments needs to transition from episodic investment (such as that provided by the IS-ENES projects) to a more sustained mode, which should also include the provision of support for the underlying standards as well as for common policies and rules. Opportunities for additional collaborative research into data standards, workflow, data handling, data documentation, and data analytic technologies should be created. In particular, new approaches based on data-intensive facilities running high-performance analytics frameworks jointly with server-side analysis capabilities or next generation distributed file system over high performance network need to be explored. Data intensive facilities (representing the counterpart to the HPC eco-system for generating simulations) close to the different storage hierarchies or having access to the different storage hierarchies will be needed to address high-performance scientific data management. In such eco-system, joining HPC and Big Data, parallel applications and in-situ frameworks for climate data analysis and diagnostics would provide a new generation of "*data tools*" for climate scientists. Finally, the climate science community needs to establish a better interface with the downstream user communities such as the climate services community to ensure that data and data service requirements meet capability.

Revised Recommendation: The ENES data infrastructure should evolve towards a sustained infrastructure, ensuring that data from climate simulations are easily available, well-documented and quality assured, for both climate scientists and downstream users. The community should further invest in research into data standards, workflow, data handling, high performance data management and data analytics to meet the challenges of increasing (big) data volumes and complexity.

Recommendation (iv): Physical network

ENES should continue to support the increase in base bandwidth for national and European networks for research. All sites should continue to monitor network performance (in general, and end-to end), particularly during the execution of significant modelling campaigns involving sustained data transfers from HPC to remote data archives. All sites should also monitor and share their local ingress/egress data both in terms of absolute values, and in terms of local capacity, so pinch-points can be avoided either by enhancing data replication, or increasing local network bandwidth (or both). To exploit replication the ENES community should ensure the wide-availability of replication software as well as tools that can support the simultaneous download of data from multiple replicates and manage the workflow associated with high-volume data transfers.

Revised recommendation: 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.

Recommendation (v): People

ENES still needs to strengthen the network of science experts and software engineers, enhance common developments, and develop training as recommended in 2012. Summer schools could be streamlined and made more efficient so that they could be delivered more frequently without always engaging the same individuals. Software engineering training needs to be prioritised, and material and best practices shared between sites. Institutions need to prize a culture that reflects the mutual dependencies between climate science, computer science, software engineering, and systems designers and administrators. The community needs to be pro-active in attracting high calibre individuals, and in making and seizing opportunities to influence undergraduate training and postgraduate topics in parallel disciplines. High calibre individuals are often stimulated by complex and hard problems - the community needs to be clearer about the scale of the challenges from production science to developing next generation codes. Technical training and careers need to be as easily available and valued as scientific training and careers.

Revised Recommendation: Grow the numbers of skilled scientists and software engineers in the ENES community, both by attracting new people and enhancing the skills of those already in the field. In doing so, be proactive about advertising the intellectual and technical challenges in climate science, 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 between software engineers across the community**.**

5.2 Additional Recommendations in 2017

Climate science faces challenges over a range of delivery scales, from production climate science exploiting the current generation of models, to developing next generation codes. Since 2012 it has become clear that there are two further recommendations needed to help the community face the scientific challenges ahead, one focusing on model evaluation, and one focusing on organisational challenges.

Recommendation (vi): Infrastructure for the scientific evaluation of models

The community should continue to address the consequences of recognizing that the infrastructure for model evaluation is both hard to develop and maintain and yet an integral and routine part of developing and evaluating models. Opportunities to increase the level of common development and sharing of standard diagnostic and evaluation tools should be taken, with the aim of supporting the acceleration of model development and analysis of the model output.

New Recommendation: The community should further enhance the sharing of common open source diagnostics and evaluation tools. Shared governance procedures for climate model evaluation should be put in place, covering both the evaluation aims and the software structures. Data infrastructure should be expanded to include the computational resources needed for running the evaluation tools on model output as they are submitted to community repositories.

Recommendation (vii): The organisational challenge for climate science infrastructure

The climate modelling community is facing major infrastructure challenges both in developing and sustaining the international collaborations needed to run and share the WCRP reference experiments and in preparing next generation of climate models. To address the first, ENES needs to further develop and sustain its data infrastructure to support European contribution to international climate science, strengthen its role in the international ESGF governance, and further enable the development of climate services. For the latter, ENES needs to continue to support European model developments across climate science, from particular problems such as improving model reliability at regional scale, to the efficient use of complex and fast evolving future computer, storage and network architectures.

New Recommendation: The ENES community should develop a sustained cooperation that put its infrastructure for both the production of international reference experiments and the development of future models on a firmer footing, utilising both national and European funding. In doing so, ENES needs to strengthen its collaboration with other European actors providing services (e.g. GEANT, PRACE) or using ENES services (e.g., the Copernicus climate change services).

6. References

Brasseur, G., and D. Carlson (2015), Future directions for the World Climate Research Programme, *Eos, 96,* 30 July 2015. doi:10.1029/2015EO033577

Eyring, V., S. Bony, G.A. Meehl, C.A., Senior, B. Stevens, R.J. Stouffer, and K.E. Taylor (2016a): Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev., 9*, 1937-1958, doi:10.5194/gmd-9-1937-2016

Eyring, V., P.J. Gleckler, C. Heinze, R. J. Stouffer, K. E. Taylor, V. Balaji, E. Guilyardi, S. Joussaume, S. Kindermann, B. N. Lawrence, G. A. Meehl, M. Righi, and Dean N. Williams (2016b), Towards improved and more routine Earth system model evaluation in CMIP, *Earth System Dynamics, 7,* 813–830, doi:10.5194/esd-7-813-2016

Marotzke J., C. Jakob, S. Bony, P. A. Dirmeyer, P. A. O'Gorman, E. Hawkins, S. Perkins-Kirkpatrick, C. Le Quéré, S. Nowicki, K. Paulavets, S. I. Seneviratne, B. Stevens, M. Tuma (2017), Climate research must sharpen its view, Commentary, *Nature Climate Change*, *7*, doi: 10.1038/nclimate3206

Meehl G. A., R. Moss, K. E. Taylor, V. Eyring, R. J. Stouffer, S. Bony, B. Stevens (2014), Climate model intercomparisons: preparing for the next phase, *Eos*, *95*, 9, pp 77-84,014.

Mitchell J., R.Budich, S. Joussaume, B. Lawrence and J. Marotzke (2012), Infrastructure strategy for the European Earth System Modelling community 2012-2022, *ENES Report Series 1,* 33pp, doi.org/10.5285ca90b281d6ff4cffb9a9bbdeb5fa63f3; also available on http://enes.org

Schulthess T. (2015), Programming revisited, *Nature Physics, 11*, 369-373. doi:10.1038/nphys3294

Street R., M. Parry, J. Scott, D. Jacob and T. Runge (2015), A European research and innovation roadmap for climate services, 51 pp, doi: 10.2777/702151