

NICEST-2

1. Significance

Towards Open Science, reproducibility of earth system runs is as important as producing accurate and Findable Accessible Interoperable and Reusable ([FAIR](#)) climate data. The EuroHPC initiative is a unique opportunity for the Nordic countries towards a joint Climate Model Intercomparison Project ([CMIP](#)) participation provided that all the necessary building blocks (containerization, workflow management, job orchestration, model diagnostics) for creating fully reproducible and efficient Earth System Model (ESM) workflows as well as for publishing FAIR climate data are in place.

NICEST-2 aims at i) improving the e-infrastructure to tackle these challenges in optimal conditions, with minimal waste of resources, and ii) setting up the building blocks for ESM workflows which are to be deployed as part of the Nordic European Open Science Cloud (EOSC), thereby realizing its full potential for the benefit of the Nordics.

Expected benefits

- More ESM simulations done in less time through performance enhancements of the codes and optimization of the workflows for the entire community
- Homogeneous workflows for ESMs (NorESM, EC-Earth)
- Common diagnostic tools/workflows for highlighting climate impact in the Nordics
- Lower cost and better energy efficiency of the simulations
- Possibility to have a joint Nordic CMIP participation
- Benefit for all ESM researchers in the Nordic, not only those involved in CMIPs
- Climate data would be easier to access and exploit, even for non-specialists

1.1 Research community

A major challenge for the international climate research community is the development of comprehensive ESMs capable of simulating natural climate variability and human-induced changes. Such models need to account for processes occurring in the atmosphere, the ocean and on the continents including physical, chemical and biological processes on a variety of spatial and temporal scales. They also have to capture complex nonlinear interactions between the different components of the earth system and assess how these interactions can be perturbed as a result of human activities.

Accurate scientific information is required at various levels of society, from governments and businesses to the general public, to make appropriate decisions regarding our global environment, with direct consequences on the economy and lifestyles. It is therefore the responsibility of the scientific community to accelerate progress towards a better understanding of the processes governing the earth system and towards the development of an improved predictive capability. In addition, it is paramount that the latest research results

are openly and freely disseminated to all stakeholders and that relevant tools are supplied for them to extract information meeting their needs.

The **Nordic climate modeling community** consists of research groups at Universities, national meteorological institutes and research institutes, and holds demonstrable world-class excellence in the field. Through international collaborations within Europe and the US, these Nordic groups contribute to the development of both global and regional climate models (RCMs) and have used these complex and computationally demanding research tools to help understand a wide range of natural and forced phenomena within the earth system. This research is by nature, multi-disciplinary and multi-scale.

1.2 National and international context

Earth system modeling and CMIP

An important way to get ESMs internationally recognized is through the World Climate Research Program ([WCRP](#)) sponsored CMIPs which is a unique international initiative that highlights the importance of sharing, comparing and analyzing the outcomes of global climate model simulations and to deliver high quality climate information. CMIP results have been used in a very large number of scientific studies that make them a cornerstone of IPCC assessment reports. **Danish, Finnish, Norwegian and Swedish modeling groups** have committed to participate in phase 6 of CMIP ([CMIP6](#)).

The regional counterpart to CMIP is the Coordinated Regional Downscaling Experiment ([CORDEX](#)) endorsed by the WCRP that has produced a large ensemble of RCM integrations at high horizontal resolution for all continents. The CORDEX results are now being used worldwide for climate impact studies and climate change adaptation work.

ENES and IS-ENES: The European Network for Earth System modeling

Several partners in this proposal are actively involved in the [ENES/IS-ENES](#) European collaboration. ENES aims to i) help in the development and evaluation of state-of-the-art climate and ESMs, ii) facilitate focused model intercomparisons in order to assess and improve them, iii) encourage exchanges of software and model results, iv) help in the development of high-performance computing facilities dedicated to long high-resolution, multi-model ensemble integrations.

The Climate Modelling community is an essential component of joint European efforts to build a European framework of earth system modelling as part of the ENES/IS-ENES network (European Network for Earth System Science), through Horizon 2020 projects (e.g. CRESCENDO), as well as Joint Programming Initiative and possibly European Strategy Forum on Research Infrastructures (ESFRI) instruments.

Infrastructure for Norwegian Earth System modelling (INES)

Norwegian partners of the NICEST-2 proposal are all engaged in the INES project (funded by the Norwegian Research Council) that aims at establishing the next generation climate model for Norwegian and Nordic research and at providing society and business with the most recent high-quality knowledge base, at a time when climate change is understood and

recognized as a primary driver for change in wide socio-economic areas of Norway. The INES project is unique in Norway, serving climate-related user communities in a cost-efficient way, which will further enhance knowledge transfer to all researchers in the Nordics.

1.3 The Nordics and NeIC

Improving the scientific accuracy of the ESMs requires huge computing and storage resources that are still provided locally, at national level. This hinders bit-replicability and makes it difficult for researchers to compare their developments and confidently validate the scientific results. While this research is essential to assess the drastic impacts of climate change in the Nordics, Nordic countries cannot expect this research to be carried out by other international teams who obviously focus on their own regional issues.

Within the Nordic ESM modeling community, there is significant and sustained support for the concept of a Nordic collaboration. Several initiatives have been initiated in response to this ongoing interest.

NICEST: the objective of the [NICEST](#) project is to strengthen the Nordic ESM community by supporting the efficient use of various e-infrastructures through competence building, sharing and exchanging knowledge. The project will end in december 2019.

NordicESM: the NordForsk [NordicESM](#) project investigated the potential modes and merits of a coordinated/integrated Nordic platform and collaboration for earth system modeling.

European Open Science Cloud (EOSC) Nordic aims at bridging e-services in the Nordic region with European Open Science Cloud (EOSC). The Nordic Climate Community is represented in EOSC-Nordic by a few partners of the NICEST project. NICEST-2 complements this unique climate science demonstrator by ensuring that i) the EOSC-Nordic services are fully adopted by the Nordic community, ii) relevant training material is developed and introduced to researchers from different disciplines, iii) other use cases can also be easily deployed on EOSC-Nordic services following the same approach.

1.4 Services and tools

EC-Earth and NorESM

In the Nordics, two major ESMs are being used, namely [EC-Earth](#) and [NorESM](#).

EC-Earth is developed as part of a Europe-wide consortium thus promoting international cooperation and access to knowledge. The EC-Earth consortium's main objective is to develop and apply an ESM for providing trustworthy information to climate services and to advance scientific knowledge on the earth system, its variability, predictability and long-term changes resulting from external forcing. EC-Earth has become a prominent state-of-the-art model within the European landscape of ESMs, as shown by the involvement in many European projects, its significant contribution to CMIP5 and the plans for CMIP6.

NorESM is being developed by a national consortium consisting of NORCE, MET, UiB, NERSC, UiO, NILU and CICERO and is being run on various HPCs in Norway and Europe. Presently, the model is used in at-least five different countries with more than 25 scientists currently active developers of the model and many more involved in analysis and applications of model diagnostics. In addition to the comprehensive experiments made available to the community through CMIP, NorESM is also used in Norway to study present and past climate states and variability ranging from seasonal to multi-centennial time scales. Climate prediction capability in Norway is being developed with NorESM as the main tool.

NICEST-2 project will contribute to accelerate developments by avoiding re-work and facilitating exchanges between partners in the Nordics, whatever the computer and storage resources used locally.

The Earth System Model eValuation Tool ([ESMValTool](#)) is a community diagnostics and performance metrics tool for the evaluation of ESMs that allows for routine comparison of single or multiple models, either against predecessor versions or against observations. The ESMValTool is a community effort open to both users and developers encouraging open exchange of diagnostic source code and evaluation results from the CMIP ensemble.

NICEST-2 project will help to widen the usage of ESMValTool within the Nordics (beyond those involved in CMIP6) and help to better estimate climate impacts by focusing on diagnostics over the Nordic countries.

The Earth System Grid Federation: climate models produce very large amounts of output data that needs to be post-processed and quality-checked before it can be distributed to climate researchers, those working with climate impacts and other users involved with climate adaptation measures. The storage, distribution and handling of such huge dataset poses major challenges to the climate modeling community as data providers. To meet the increasing needs from the climate modeling community to archive and distribute climate data the Earth System Grid Federation ([ESGF](#)) was launched in the US in the late 1990s. CMIP and CORDEX now use the ESGF for distribution of climate data. With the support from the NeIC NICEST project, Danish, Norwegian, Finnish and Swedish ESGF nodes have been set up and are now starting to publish CMIP6 data.

Through its first phases, the ESGF grew into a Petabyte facility and will continue to increase further with the next phases of CMIP and CORDEX. The ESGF system is the main infrastructure for archiving and distributing climate data for the foreseeable future and future developments are expected to support this system. NICEST-2 project will reinforce the ESGF collaboration within the Nordics and fully realize the potential of EOSC-Nordic.

2. Activities

The project will meet its objectives with three typologies of mutually linked activities:

- **Activity A:** Model analysis and diagnostics focused on the Nordic regions
- **Activity B:** FAIR climate data for NorESM and EC-Earth
- **Activity C:** ESM workflows to efficiently run NorESM and EC-Earth on euroHPC

The activities of the project are described in terms of Tasks; each producing an output that contributes to either a project deliverable or to other Tasks (as input). Activity lead is in **bold**.

Activity A: Model analysis and diagnostics useful for the Nordic collaboration

(Partners: **MetNo**, FMI, DMI, NERSC, UiO, SMHI, INA)

The main objective of this activity is to develop expertise on ESMValTool within the Nordic countries and enable the development of new diagnostic modules.

Three main tasks have been identified:

- **Task A-1 (MetNo, NERSC)**: Organization of workshop/hackathon at the beginning of the project on ESMValTool to collect user requirements (*M1-M6*)
- **Task-A2 (FMI, MetNo, NERSC)**: Development of new ESMValTool diagnostic modules for the Nordic regions (*M7-M24*)
- **Task-A3 (NERSC, MetNo, FMI, UiO)**: Development of teaching material for learning how to use ESMValTool diagnostics over the Nordic regions (*M18-M36*)

Milestones and deliveries

- **OA-1**: 1 workshop/hackathon to discuss about how to use ESMValTool with a focus on the Nordic countries and collect user requirements for the two training materials.
- **OA-2**: Development of new ESMValTool diagnostic modules for the Nordic regions (eg, polar lows, European and Greenland blocking events) with strong emphasis on (high-latitude) observations (and "observation simulators" to some extent)
- **OA-3**: 1 online tutorial on how to best use ESMValTool diagnostic tools for the Nordic countries (eg, polar lows, European and Greenland blocking events) with emphasis on (high-latitude) observations (and "observation simulators" to some extent).

Activity B: FAIR Climate data for Nordics (Partners: **NSC**, DMI, CSC, NORCE, MetNo)

The main objective of this activity is to define a clear roadmap to FAIR climate modeling data in the Nordics. This activity will be strongly linked to EOSC-Nordic, in particular WP4 "FAIR data" and WP5 "Open Research demonstrators". Two main tasks have been identified:

- **Task B-1 (NSC, CSC, NORCE): Support for Nordic ESGF hosting of CMIP6 data** (*M1-M36*)
Danish, Finnish, Norwegian and Swedish HPC providers currently host ESGF nodes. Although there have been considerable improvements in the ESGF software, the deployment and reliable maintenance of an ESGF node requires experience and a technical understanding of the hardware and ESGF software. In addition, knowledge of climate model output formats, project data reference syntax and controlled vocabularies (e.g. CMIP6) is also required. This activity will continue to strengthen the collaborative efforts, knowledge sharing and skills development initiated in the NICEST project regarding the deployment and administration of ESGF nodes.
- **Task B-2 (NSC, NORCE, MetNo) : Towards FAIR Nordic climate modeling data**
Beyond CMIP6 and CORDEX, there are Nordic and national projects that generate valuable climate model data. The management of these outputs should follow FAIR

and open data principles. The aim of this sub-activity is to investigate to what extent the tools (including ES-DOC, CF, CMOR, ESGF) and techniques for managing CMIP6 data can be reused in Nordic and national projects to ensure the valuable outputs from such projects are FAIR and open. This will be achieved by bringing together data generators (ESM modelers), data managers (ESGF) and FAIR data specialists (*M4-M15 and M31-M36*).

Milestones and deliveries

- **OB-1.1:** Online minutes from all Nordic ESGF operation meetings (e.g., one per month for all the duration of the project)
- **OB-1.2:** 1 face to face workshop for Nordic ESGF operations team on the ESGF CMIP6 publication process.
- **OB-2.1:** 1 hackathon with FAIR experts and ESM specialists to understand what needs to be done for making Climate data FAIR (for NorESM and EC-Earth).
- **OB-2.2:** Report on knowledge/skills gap analysis and suggested roadmap to FAIR Nordic climate modeling data
- **OB-2.3:** 1 Data Management Plan Template, including best practices for FAIR Nordic Climate Data
- **OB-2.4:** 1 Data stewardship workshop for FAIR climate data

Activity C: ESM workflows for efficiently running both NorESM and EC-Earth on EuroHPC (*Partners: UiO, NERSC, NORCE, DMI, FMI*)

The main objective of this activity is to develop expertise on how to run efficiently ESM workflows and facilitate the porting and deployment of Nordic ESMs on future HPCs, in particular the EuroHPC. Two main tasks have been identified:

Task C-1 (UiO, NORCE, NERSC, FMI): Reproducible workflows for deploying and running ESM on future HPCs (*M4-M9 and M25-M30*)

As part of this task an analysis of existing workflow Management systems (cylc, galaxy, snakemake, etc.) will be done and will be the basis for the definition of Nordic ESM workflow Management Systems. Mock-up ESM workflows will be made available by mid-term and will facilitate benchmarking of Nordic ESMs on new HPCs and ease the deployment of Nordic ESMs on EuroHPCs. To improve reproducibility, package management system and environment management system such as conda will be tested, and containers for running these models will be deployed. The performance achieved by the containerized models will be made publicly available.

Within High Energy Physics (e.g., CERN LHC), many solutions and services do already exist to tackle complex tasks like workflow management execution with data and compute. Those solutions are being used now by more and more scientific communities. One potential idea is to identify some use cases and evaluate the services available like job management, data catalog and orchestration. A demonstrator will be set up as part of the project with technical support and expertise from **NORDIC WLCG TIER-1 FACILITY**.

Task C-2 (NORCE, UiO, FMI): Efficiency of Nordic ESMs on future euroHPC (*M13-M24 and M34-M36*)

Having a common framework to analyze the performance of Nordic ESMs can considerably reduce the cost (computing and energy efficiency) for running climate models and give us a competitive advantage for running on the future computing platforms, including EuroHPC. This task will focus on best practices for performance and productivity of Nordic ESMs and on understanding what we need to do to be able to run on EuroHPC.

Milestones and deliveries

- **OC-1.1:** 1 deliverable/report on the identification of the Nordic ESM community needs for ESM workflows (analysis of existing tools such as cylc, galaxy, snakemake) and identification/classification of associated services to be deployed for efficiently orchestrating these workflows on various computing platforms.
- **OC-1.2:** 1 demonstrator build with a simple use case for distributing computations and managing data on at least 3 different sites (within the Nordic countries). The associated deliverable will be publicly available.
- **OC-1.3:** 1 online training material on best practices for the deployment of ESM workflows (package management, writing efficient workflows to optimize computing and storage resources)
- **OC-2.1:** 1 workshop (face to face meeting) with national HPC providers to discuss and characterize the nordic ESM involvement in EuroHPC
- **OC-2.2:** 1 deliverable/report on bottlenecks that would hindrance the efficient usage of Nordic ESMs on EuroHPC and possible remediation actions (I/Os, adding GPU support, etc.) with a clear information on costs (PMs, etc.)
- **OC-2.3:** Best practice guides on how to evaluate the performance of Nordic ESMs.

3. Resources and viability of the anticipated e-infrastructure services

3.1 Staff cost

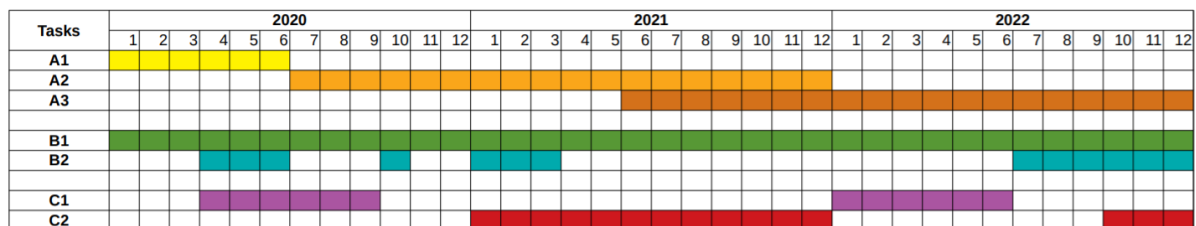
This project will operate on a budget of about 2.6 FTE/year and a total of 7.8 FTE with the following composition (*All numbers are in person-months*):

	2020	2021	2022	<i>Total</i>
MetNo	1	2	2	5
NSC	2	3	3	8
NORCE	1	1.5	1.5	4
UiO	2	3	3	8
SMHI	0	0	0	0
FMI	1	3	2	6
DMI	0	0	0	0

CSC	1	1	1	3
INAR (UoH)	0	0	0	0
NERSC	0.5	0.5	3	4
NeIC funding				38
Matching funding				38
Management	6	6	6	18
Total				94

Average cost for 1 FTE: NOK 1 600 000

3.2 Gantt chart



3.3 Travel and other budget

	2020	2021	2022
Workshops/meetings	165 kNOK	165 kNOK	165 kNOK
Attending workshops/meetings	58 kNOK	58 kNOK	58 kNOK

Sustainability plan:

The project will produce a bundle of services component supporting the entire research lifecycle management: running climate models in a reproducible way on future EuroHPC, national HPCs and cloud computing resources (containerization), data discovery and access to processing (job orchestration), dissemination (ESGF), result validation, reuse and executable paper publishing for peer review in accordance with open journal data policies and e-learning training material. Those services can be procured and operated either “à la Carte” or through a Services bundle via EOSC-Nordic, for the specific need and user segment – Nordic Climate as a Service - according to sustainability business models to be developed within the project and EOSC-Nordic. Co-design is an integral part of our approach and several workshops will be organized to ensure the long-term sustainability.

Partners of the collaboration

The **National Supercomputer Centre in Sweden (NSC)** provides leading edge high performance computing resources and support to users throughout Sweden. NSC is an independent center within Linköping University and is funded by the Swedish Research Council via SNIC. NSC also has a long established collaboration with SMHI, delivering cluster system services, storage solutions, data publication services and advanced user support. NSC acts as host and technical administrator of the SMHI-NSC [ESGF data node](#) and provides ESM technical support for SMHI and Swedish academic climate modeling groups. NSC is working closely with the Rossby Centre at SMHI, and the [Bolin Centre](#) for Climate Research assisting with the coordination and resource planning for the Swedish contribution to the CMIP6 international climate modeling project.

CSC – IT Center for Science Ltd. provides IT support and modeling, computing and information services for academia, research institutes and companies in Finland. CSC manages the Finnish national research and education network (FUNET) that provides fast internet access to universities and other academic institutions. CSC is part of national CoE for atmospheric science and hosts databases of national SMEAR atmosphere observation stations as well as collaborates with Finnish ESM community. CSC is the hosting partner of one of the EuroHPC pre-exascale systems.

The **Nansen Environmental and Remote Sensing Center (NERSC)** is a national environmental research institute with basic funding from the Norwegian Ministry of Climate and Environment. Modelling and prediction of current and future climate with ESM and advanced data assimilation techniques are two of the major foci at the Centre. The Centre contributes to the development of the NorESM for many years and has been leading the development of the Norwegian Climate Prediction Model (NorCPM) in recent years.

NORCE Norwegian Research Centre staff has had a prominent role in the development and application of NorESM since 2007. NORCE is coordinating INES, bringing together 6 national partners and aiming to upgrade and maintain a cutting-edge and verified ESM and provide an infrastructure for efficient model simulations, storage, analysis, validation and sharing of data. NORCE is strongly involved in the Norwegian CMIP6 contribution.

METNO is the national meteorological service in Norway. In addition to national and aviation weather forecasting, and climate monitoring for Norway and adjacent sea areas, the institute represents Norway in ECMWF, EUMETSAT, EUMETNET, WMO and other international fora, and takes part in national and international research projects (e.g. EU) on climate, atmospheric and marine research, and air pollution. METNO is, together with NORCE Climate and the universities in Oslo and Bergen, the main developer of the NorESM, providing results for CMIP and IPCC.

Finnish Meteorological Institute's (FMI) mission is to produce high-quality observation and research data on the atmosphere and seas. The Institute combines this expertise into the services it provides to promote public safety as well as increase well-being in the society and the environment, taking into account the needs to maintain preparedness.

Danish Meteorological Institute (DMI) participates in a large number of research and development projects, which are highly interdisciplinary and linked with many other institutions nationally and internationally. Activities are presented thematically and heavily cross-linked below in order to provide easy access to people, projects, publications or possible collaboration opportunities.

The University of Oslo (UiO): the **Meteorology and Oceanography section (MetOs)** is involved in several national and European projects (INES, KeyClim, VIKINGS, EMERALD, etc.) related to Earth System Modeling. The **Department for Research Infrastructure Services** of University Center for Information Technology (**USIT**) will participate in task C-1. **Research Infrastructure Services** Group is involved in NORDIC WLCG TIER-1 facility.

Swedish Meteorological and Hydrological Institute (SMHI) is an expert agency under the Ministry of the Environment and Energy. Through unique expertise in meteorology, hydrology, oceanography and climatology, SMHI contributes towards greater public welfare, increased safety and a sustainable society.

Institute for Atmospheric and Earth System Research at University of Helsinki (UHEL-INAR) is a multi- and interdisciplinary research unit that strengthens the internationally leading, integrated multidisciplinary research and education environment for atmospheric and earth system science and to feed in scientific results for the national and international environment and climate policy.