Please check our wiki for help on navigating the form.

EuroHPC-IA

Call: H2020-JTI-EuroHPC-2019-1

(Innovating and Widening the HPC use and skills base)

Topic: EuroHPC-02-2019

Type of action: EuroHPC-IA

Proposal number: 955558

Proposal acronym: eFlows4HPC

Deadline Id: H2020-JTI-EuroHPC-2019-1

Table of contents

How to fill in the forms

The administrative forms must be filled in for each proposal using the templates available in the submission system. Some data fields in the administrative forms are pre-filled based on the steps in the submission wizard.

1 - General information

Topic	EuroHPC-02-2019	Type of Action	EuroHPC-IA
Call Identifier	H2020-JTI-EuroHPC-2019-1	Deadline Id	H2020-JTI-EuroHPC-2019-1

Acronym	eFlows4HPC
Proposal title	Enabling dynamic and Intelligent workflows in the future EuroHPCecosystem

Note that for technical reasons, the following characters are not accepted in the Proposal Title and will be removed: < > " &

Duration in months	*36*
Fixed keyword 1	Artificial intelligence, intelligent systems, multi agent systems
Fixed keyword 2	High performance computing
Fixed keyword 3	Machine learning, statistical data processing and applications using
Fixed keyword 4	Numerical analysis, simulation, optimisation, modelling tools
Free keywords	*Scientific and Industrial Workflows, Convergence High-Performance Data Analytics, Machine Learning and High-Performance Computing, Distributed and heterogeneous computing, Widening HPC usage*

*Abstract**

Today, developers lack tools that enable the development of complex workflows involving HPC simulation and modelling with data analytics (DA) and machine learning (ML). TheFlows4HPC aims to deliver a workflow software stack and an additional set of services to enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The software stack will allow to develop innovative adaptive workflows that efficiently use the computing resources and also considering innovative storage solutions.
To widen the access to HPC to newcomers, the project will provide HPC Workflows as a Service (HPCWaaS), an environment for sharing, reusing, deploying and executing existing workflows on HPC systems. The workflow technologies, associated machine learning and big data libraries used in the project leverages previous open source European initiatives. Specific optimization tasks for the use of accelerators (FPGAs, GPUs) and the EPI will be performed in the project use cases.
To demonstrate the workflow software stack, use cases from three thematic pillars have been selected. Pillar I focuses on the construction of DigitalTwins for the prototyping of complex manufactured objects integrating state-of-the-art adaptive solvers with machine learning and data-mining, contributing to the Industry 4.0 vision. Pillar II develops innovative adaptive workflows for climate and for the study of Tropical Cyclones (TC) in the context of the CMIP6 experiment, including in-situ analytics. Pillar III explores the modelling of natural catastrophes - in particular, earthquakes and their associated tsunamis- shortly after such an event is recorded. Leveraging two existing workflows, the Pillar will work of integrating them with the eFlows4HPC software stack and on producing policies for urgent access to supercomputers. The pillar results will be demonstrated in the target community CoEs to foster adoption and get feedback.

Remaining characters 2

Has this proposal (or a very similar one) been submitted in the past 2 years in response to a call for proposals under Horizon 2020 or any other EU programme(s)? ○ Yes ◉ No

Please give the proposal reference or contract number.	
xxxxxx-x	

Declarations

1*) We/I declare to have the explicit consent of all participants on their participation and on the content of this proposal.	☒
2) We/I confirm that the information contained in this proposal is correct and complete and that none of the project activities have started before the proposal was submitted.	☒
3) We/I declare: - to be fully compliant with the eligibility criteria set out in the call - not to be subject to any exclusion grounds under the EU Financial Regulation 2018/1046 - to have the financial and operational capacity to carry out the proposed project.	☒
4) We/acknowledge that all communication will be made through the Funding & Tenders Portal electronic exchange system and that access and use of this system is subject to the Funding & Tenders Portal Terms and Conditions.	☒
5) We/I acknowledge and authorize the collection, use and processing of personal data for the purpose of the evaluation of the proposal and the subsequent management of the grant/prize (if any). We/I acknowledge and authorize that the data may also be used for the monitoring and evaluation of the EU funding programme, the design of future programmes and communication purposes.	☒
6) We/I declare that subcontracts will be best value for money and free of conflict of interest.	☒
7*) We/I declare that all beneficiaries have followed their own accounting practices for the preparation of the budget and have included therein only cost that would be eligible for an actual costs grant, excluding costs that are ineligible under H2020 rules.	☒

The coordinator is only responsible for the correctness of the information relating to his/her own organisation. Each applicant remains responsible for the correctness of the information related to him/her and declared above. If the proposal to be retained for EU funding, the coordinator and each beneficiary will be required to present a formal declaration in this respect.

Note:

For **multi-beneficiary applications**, the coordinator vouches for its own organization and that all other participants confirmed their participation and compliance with conditions set out in the call. If the proposal is retained for funding, each participant will be required to submit a formal declaration of honour confirming this.

False statements or incorrect information may lead to administrative sanctions under the Financial Regulation 2018/1046.

Personal data will be collected, used and processed in accordance with Regulation 2018/1725 and the **Funding & Tenders Portal privacy statement**.

Please be however aware that, to protect EU financial interests, your data may be transferred to other EU institutions and bodies and be registered in the EDES database. Data in the EDES database is also subject to Regulation 2018/1725 and the EDES privacy statement.

2 - Participants & contacts

#	Participant Legal Name	Country	Action
1	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	ES	
2	CENTRE INTERNACIONAL DE METODES NUMERICS EN ENGINYERIA	ES	
3	FORSCHUNGSZENTRUM JULICH GMBH	DE	
4	UNIVERSITAT POLITECNICA DE VALENCIA	ES	
5	BULL SAS	FR	
6	DtoK Lab S.r.l.	IT	
7	FONDAZIONE CENTRO EURO-MEDITERRANEOSUI CAMBIAMENTI CLIMATICI	IT	
8	INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE	FR	
9	SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI DI TRIESTE	IT	
10	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK	PL	
11	UNIVERSIDAD DE MALAGA	ES	
12	ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA	Italy	
13	ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG	DE	
14	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	CH	
15	SIEMENS AKTIENGESELLSCHAFT	DE	
16	STIFTELSEN NORGES GEOTEKNISKE INSTITUTT	NO	

2 - Administrative data of participating organisations

PIC	***Legal name***
999655520	*BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION*

Short name: BSC

Address

Street Calle Jordi Girona 31

Town BARCELONA

Postcode 08034

Country Spain

Webpage www.bsc.es

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....01/03/2005 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Computer Sciences ☐ not applicable

☒ Same as proposing organisation's address

Street: Calle Jordi Girona 31

Town: BARCELONA

Postcode: 08034

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: ○ Male ◉ Female

First name: **Rosa M** Last name: **Badia**

E-Mail: **rosa.m.badia@bsc.es**

Position in org.: WORKFLOWS AND DISTRIBUTED COMPUTING GROUP MANAGER

Department: Computer Sciences ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: Calle Jordi Girona 31

Town: BARCELONA Post code: 08034

Country: Spain

Website: www.bsc.es

Phone: +xxx xxxxxxxxx Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	*Last Name*	*E-mail*	*Phone*
Vanessa	Fernandez	vanessa.fernandez@bsc.es	+xxx xxxxxxxxx

PIC	***Legal name***
999658721	*CENTRE INTERNACIONAL DE METODES NUMERICS EN ENGINYERIA*

Short name: CIMNE

Address

Street C GRAN CAPITAN, EDIFICI C1, CAMPUS NOR

Town BARCELONA

Postcode 08034

Country Spain

Webpage www.cimne.com

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....19/08/2008 - no

SME self-assessment ..... unknown

SME validation sme.....19/08/2008 - no

Department(s) carrying out the proposed work

Department 1

Department name: KRATOS Group ☐ not applicable

☒ Same as proposing organisation's address

Street: C GRAN CAPITAN, EDIFICI C1, CAMPUS NORD

Town: BARCELONA

Postcode: 08034

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Riccardo** Last name: **Rossi**

E-Mail: **rrossi@cimne.upc.edu**

Position in org.: Full Research Professor

Department: Kratos Group ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: C GRAN CAPITAN, EDIFICI C1, CAMPUS NORD UPC SN

Town: BARCELONA Post code: 08034

Country: Spain

Website: https://www.cimne.com/verperfil/4803

Phone: +34 93 401 56 96 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Marina	De la Cruz	projectes@cimne.upc.edu	*+34 93 401 74 95*
Cecilia	Soriano	csoriano@cimne.upc.edu	*+xxx xxxxxxxxx*

PIC	***Legal name***
999980470	*FORSCHUNGSZENTRUM JULICH GMBH*

Short name: FZJ

Address

Street WILHELM JOHNEN STRASSE

Town JULICH

Postcode 52428

Country Germany

Webpage www.fz-juelich.de

Specific Legal Statuses

Legal person .....yes

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....05/12/1967 - no

SME self-assessment ..... unknown

SME validation sme.....05/12/1967 - no

Department(s) carrying out the proposed work

Department 1

Department name: Juelich Supercomputing Centre (JSC) ☐ not applicable

☒ Same as proposing organisation's address

Street: WILHELM JOHNEN STRASSE

Town: JULICH

Postcode: 52428

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **Daniel** Last name: **Mallmann**

E-Mail: **d.mallmann@fz-juelich.de**

Position in org.: Scientist in Charge

Department: Juelich Supercomputing Centre (JSC) ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: WILHELM JOHNEN STRASSE

Town: JULICH Post code: 52428

Country: Germany

Website: www.fz-juelich.de

Phone: +49 2461 612433 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Volker	Marx	v.marx@fz-juelich.de	*+49 2461 615831*

PIC	***Legal name***
999864846	*UNIVERSITAT POLITECNICA DE VALENCIA*

Short name: UPV

Address

Street CAMINO DE VERA SN EDIFICIO 3A

Town VALENCIA

Postcode 46022

Country Spain

Webpage www.upv.es

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....26/03/1971 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Enrique S. Quintana-Orti ☐ not applicable

☐ Same as proposing organisation's address

Street: Camino de Vera, s/n

Town: Valencia

Postcode: 46022

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Enrique S.** Last name: **Quintana-Orti**

E-Mail: **quintana@disca.upv.es**

Position in org.: Professor

Department: Informática de Sistemas y Computadores ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Camino de Vera, s/n

Town: Valencia Post code: 46022

Country: Spain

Website:

Phone: +34963877000 Phone 2: +34680301814 Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Carles	Hernandez	carherlu@upv.es	*+34963877000*
Jose	Flich-Cardó	jflich@disca.upv.es	*+34963877000*
Carmen	Navarro Tomás	cnavarro@gap.upv.es	*+34963877000*

PIC	***Legal name***
996058081	*BULL SAS*

Short name: BULL

Address

Street RUE JEAN JAURES 68

Town LES CLAYES SOUS BOIS

Postcode 78340

Country France

Webpage www.bull.com

Specific Legal Statuses

Legal person .....yes

Public body .....no

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Industry (private for profit).....yes

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....08/11/2017 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Big Data and Security - R&D ☐ not applicable

☒ Same as proposing organisation's address

Street: RUE JEAN JAURES 68

Town: LES CLAYES SOUS BOIS

Postcode: 78340

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **Etienne** Last name: **Walter**

E-Mail: **etienne.walter@atos.net**

Position in org.: R&D Project Manager

Department: Big Data and Security - R&D ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: RUE JEAN JAURES 68

Town: LES CLAYES SOUS BOIS Post code: 78340

Country: France

Website: https://atos.net/en/products

Phone: +33130807496 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
François	Exertier	francois.exertier@atos.net	+xxx xxxxxxxxx
Claire	Chen	claire.chen@atos.net	+33130803356
Medur	Sridharan	medur.sridharan@bull.net	+33130803024

PIC	***Legal name***
915865950	*DtoK Lab S.r.l.*

Short name: DtoK Lab S.r.l.

Address

Street Piazza Vermicelli c/o Technest

Town Rende (CS)

Postcode 87036

Country Italy

Webpage www.dtoklab.com

Specific Legal Statuses

Legal person .....yes

Public body .....unknown

Non-profit .....unknown

International organisation .....unknown

International organisation of European interest .....unknown

Secondary or Higher education establishment .....unknown

Research organisation .....unknown

Industry (private for profit).....unknown

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Research and Development ☐ not applicable

☒ Same as proposing organisation's address

Street: Piazza Vermicelli c/o Technest

Town: Rende (CS)

Postcode: 87036

Country: Italy

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof.

Sex: (•) Male ( ) Female

First name: **Domenico**

Last name: **Talia**

E-Mail: **talia@dtoklab.com**

Position in org.: Head of Research and Development

Department: Research and Development

☐ Same as organisation name

☒ Same as proposing organisation's address

Street: Piazza Vermicelli c/o Technest

Town: Rende (CS)

Post code: 87036

Country: Italy

Website: www.dtoklab.com

Phone: +39 3204204970

Phone 2: *+xxx xxxxxxxxx*

Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Paolo	Trunfio	trunfio@dtoklab.com	*+39 3204238234*

PIC	***Legal name***
999419422	*FONDAZIONE CENTRO EURO-MEDITERRANEOSUI CAMBIAMENTI CLIMATICI*

Short name: FONDAZIONE CMCC

Address

Street VIA A IMPERATORE 16

Town LECCE

Postcode 73100

Country Italy

Webpage www.cmcc.it

Specific Legal Statuses

Legal person .....yes

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....11/05/2005 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: ASC - Advanced Scientific Computing ☐ not applicable

☒ Same as proposing organisation's address

Street: VIA A IMPERATORE 16

Town: LECCE

Postcode: 73100

Country: Italy

Department 2

Department name: CSP - Climate Simulations and Predictions ☐ not applicable

☐ Same as proposing organisation's address

Street: Via Carlo Berti Pichat 6/2

Town: Bologna

Postcode: 40127

Country: Italy

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Sandro** Last name: **Fiore**

E-Mail: **sandro.fiore@cmcc.it**

Position in org.: Senior Researcher

Department: ASC - Advanced Scientific Computing ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: VIA A IMPERATORE 16

Town: LECCE Post code: 73100

Country: Italy

Website: https://www.cmcc.it/people/fiore-sandro

Phone: +390832671081 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Giulia	Galluccio	giulia.galluccio@cmcc.it	*+390283623433*
Giovanni	Aloisio	giovanni.aloisio@cmcc.it	*+390832671081*
Alessandro	D'Anca	alessandro.danca@cmcc.it	*+390832671081*
Laura	Conte	laura.conte@cmcc.it	*+390832671081*
Enrico	Scoccimarro	enrico.scoccimarro@cmcc.it	*+390510301609*
Silvio	Gualdi	silvio.gualdi@cmcc.it	*+390510301605*

PIC	***Legal name***
999547074	*INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE*

Short name: INRIA

Address

Street DOMAINE DE VOLUCEAU ROCQUENCOURT

Town LE CHESNAY CEDEX

Postcode 78153

Country France

Webpage www.inria.fr

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: CARDAMOM Research Group / Inria Bordeaux-Sud-Ouest ☐ not applicable

☐ Same as proposing organisation's address

Street: 200 avenue de la vieille tour

Town: Talence

Postcode: 33405

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Mario** Last name: **Ricchiuto**

E-Mail: **mario.ricchiuto@inria.fr**

Position in org.: CARDAMOM Research Group Team Leader

Department: CARDAMOM Research Group / Inria Bordeaux-Sud-Ouest ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: 200 avenue de la vieille tour

Town: Talence Post code: 33405

Country: France

Website:

Phone: +xxx xxxxxxxxx Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Nicolas	Jahier	nicolas.jahier@inria.fr	*+33 5 24 57 41 70*
Lucia	Marta	lucia.marta@inria.fr	*+xxx xxxxxxxxx*

PIC	***Legal name***
999850878	*SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI DI TRIESTE*

Short name: SISSA

Address

Street VIA BONOMEA 265

Town TRIESTE

Postcode 34136

Country Italy

Webpage www.sissa.it

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....06/03/1978 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Mathematics Area, SISSA mathLab ☐ not applicable

☒ Same as proposing organisation's address

Street: VIA BONOMEA 265

Town: TRIESTE

Postcode: 34136

Country: Italy

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Gianluigi** Last name: **Rozza**

E-Mail: **gianluigi.rozza@sissa.it**

Position in org.: Full professor, director's delegate, Phd programme coordinator

Department: SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI DI TRIEST ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: VIA BONOMEA 265

Town: TRIESTE Post code: 34136

Country: Italy

Website: mathlab.sissa.it

Phone: +390403787451 Phone 2: +393356704403 Fax: +390403787528

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Gabriele	Rizzetto	direzioneamministrativa@sissa.it	*+390403787453*

PIC	***Legal name***
999586359	*INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK*

Short name: PSNC

Address

Street NOSKOWSKIEGO 12-14

Town POZNAN

Postcode 61 704

Country Poland

Webpage

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....11/12/1998 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Computer Sciences ☐ not applicable

☒ Same as proposing organisation's address

Street: NOSKOWSKIEGO 12-14

Town: POZNAN

Postcode: 61 704

Country: Poland

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Norbert** Last name: **Meyer**

E-Mail: **meyer@man.poznan.pl**

Position in org.: Supercomputing Department Manager

Department: Supercomputing Division ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Jana Pawla II 10

Town: Poznan Post code: 61-139

Country: Poland

Website: www.pcss.pl

Phone: +48618582001 Phone 2: *+xxx xxxxxxxxx* Fax: +48618525954

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Malgorzata	Klimas	malgos@man.poznan.pl	*+xxx xxxxxxxxx*

PIC	***Legal name***
999898311	*UNIVERSIDAD DE MALAGA*

Short name: UMA

Address

Street AVDA CERVANTES, NUM. 2

Town MALAGA

Postcode 29016

Country Spain

Webpage www.uma.es

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....19/02/2009 - no

SME self-assessment ..... unknown

SME validation sme.....19/02/2009 - no

Department(s) carrying out the proposed work

Department 1

Department name: Computer Sciences ☐ not applicable

☒ Same as proposing organisation's address

Street: AVDA CERVANTES, NUM. 2

Town: MALAGA

Postcode: 29016

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Jorge** Last name: **Macias**

E-Mail: **jmacias@uma.es**

Position in org.: Associate Professor

Department: UNIVERSIDAD DE MALAGA ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: AVDA CERVANTES, NUM. 2

Town: MALAGA Post code: 29016

Country: Spain

Website: www.uma.es

Phone: (+34)952 13 20 16 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Yamina	Seamari	ope@uma.es	*+34 952 13 41 90*
Coral	Erades	otri@uma.es	*+34 952 13 72 15*

PIC	***Legal name***
999472675	*ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA*

Short name: ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

Address

Street Via di Vigna Murata 605

Town ROMA

Postcode 00143

Country Italy

Webpage www.ingv.it

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....05/05/2016 - no

SME self-assessment .....05/05/2016 - no

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Sezione di Bologna ☐ not applicable

☐ Same as proposing organisation's address

Street: via D. Creti, 12

Town: Bologna

Postcode: 40128

Country: Italy

Department 2

Department name: Sezione di Roma 1 ☐ not applicable

☒ Same as proposing organisation's address

Street: Via di Vigna Murata 605

Town: ROMA

Postcode: 00143

Country: Italy

Department 3

Department name: Osservatorio Nazionale Terremoti ☐ not applicable

☒ Same as proposing organisation's address

Street: Via di Vigna Murata 605

Town: ROMA

Postcode: 00143

Country: Italy

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Jacopo** Last name: **Selva**

E-Mail: **jacopo.selva@ingv.it**

Position in org.: Reseacher

Department: Sezione di Bologna ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: via D. Creti, 12

Town: Bologna Post code: 40128

Country: Italy

Website: www.ingv.it

Phone: +39 0514151473 Phone 2: +39 3470970619 Fax: +39 0514151498

Other contact persons

First Name	*Last Name*	*E-mail*	*Phone*
Sonia	Vivanelli	sonia.vivanelli@ingv.it	*+39 0514151466*

PIC	***Legal name***
999497507	*ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG*

Short name: AWI

Address

Street AM HANDELSHAFEN 12

Town BREMERHAVEN

Postcode 27570

Country Germany

Webpage www.awi.de

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....31/12/2015 - no

SME self-assessment .....17/03/1986 - no

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Climate Dynamics ☐ not applicable

☒ Same as proposing organisation's address

Street: AM HANDELSHAFEN 12

Town: BREMERHAVEN

Postcode: 27570

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Thomas** Last name: **Jung**

E-Mail: **thomas.jung@awi.de**

Position in org.: Head of Climate Dynamics department

Department: Climate Dynamics ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: AM HANDELSHAFEN 12

Town: BREMERHAVEN Post code: 27570

Country: Germany

Website: https://www.awi.de/ueber-uns/organisation/mitarbeiter/thomas-jung

Phone: +49 47148311761 Phone 2: *+xxx xxxxxxxxx* Fax: +49 47148311797

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Nikolay	Koldunov	nikolay.koldunov@awi.de	*+xxx xxxxxxxxx*

PIC	***Legal name***
999979015	*EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH*

Short name: ETH Zürich

Address

Street Raemistrasse 101

Town ZUERICH

Postcode 8092

Country Switzerland

Webpage www.ethz.ch

Specific Legal Statuses

Legal person .....yes

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....06/01/2009 - no

SME self-assessment ..... unknown

SME validation sme.....06/01/2009 - no

Department(s) carrying out the proposed work

Department 1

Department name: Computer Sciences ☐ not applicable

☒ Same as proposing organisation's address

Street: Raemistrasse 101

Town: ZUERICH

Postcode: 8092

Country: Switzerland

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Andreas** Last name: **Fichtner**

E-Mail: **andreas.fichtner@erdw.ethz.ch**

Position in org.: Professor

Department: Institute of Geophysics ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Sonneggstrasse 5

Town: Zurich Post code: 8092

Country: Switzerland

Website: https://cos.ethz.ch/

Phone: +41446322597 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Agatha	Keller	grants@sl.ethz.ch	*+41446345350*

PIC	***Legal name***
999987260	*SIEMENS AKTIENGESELLSCHAFT*

Short name: SIEMENS AKTIENGESELLSCHAFT

Address

Street WERNER-VON-SIEMENS-STR. 1

Town MUNCHEN

Postcode 80333

Country Germany

Webpage www.siemens.com

Specific Legal Statuses

Legal person .....yes

Public body .....no

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Industry (private for profit).....yes

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....31/12/2013 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Department(s) carrying out the proposed work

Department 1

Department name: Corporate Technology ☐ not applicable

☐ Same as proposing organisation's address

Street: Otto-Hahn-Ring 6

Town: München

Postcode: 81739

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Stefan** Last name: **Boschert**

E-Mail: **stefan.boschert@siemens.com**

Position in org.: Senior Key Expert

Department: CT RDA SDT ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Otto-Hahn-Ring 6

Town: München Post code: 81739

Country: Germany

Website:

Phone: +49 152 22593299 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	*Last Name*	*E-mail*	*Phone*
Caroline	Wagner-Winter	caroline.wagner-winter@siemens.com	*+49 89 636 633412*

PIC	***Legal name***
998866522	*STIFTELSEN NORGES GEOTEKNISKE INSTITUTT*

Short name: NGI

Address

Street Sognsveien 72

Town OSLO

Postcode N-0855

Country Norway

Webpage www.ngi.no

Specific Legal Statuses

Legal person .....yes

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Industry (private for profit).....no

Enterprise Data

Based on the below details from the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

SME self-declared status.....31/12/2016 - no

SME self-assessment .....31/12/2016 - no

SME validation sme.....08/01/2009 - yes

Department(s) carrying out the proposed work

Department 1

Department name: Offshore Geohazards and Dynamics ☐ not applicable

☒ Same as proposing organisation's address

Street: Sognsveien 72

Town: OSLO

Postcode: N-0855

Country: Norway

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Finn** Last name: **Lovholt**

E-Mail: **finn.lovholt@ngi.no**

Position in org.: Principal Scientist

Department: Offshore Geohazards and Dynamics ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: Sognsveien 72

Town: OSLO Post code: N-0855

Country: Norway

Website: www.ngi.no

Phone: +47 95 79 31 00 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Thomas	Langford	thomas.langford@ngi.no	*+47 478 95 755*

3 - Budget

No	Participant	Country	(A) Direct personnel costs/€	(B) Other direct costs/€	(C) Direct costs of sub-contracting/€	(D) Direct costs of providing financial support to third parties/€	(E) Costs of inkind contributions not used on the beneficiary's premises/€	(F) Indirect Costs / € (=0.25(A+B-E))	(G) Special unit costs covering direct & indirect costs / €	(H) Total estimated eligible costs / € (=A+B+C+D+F+G) BENEFICIARY	(I) Reimburse-ment rate (%) BENEFICIARY	(J) Max.EU Contribution / € (=H*I) BENEFICIARY	(K) Costs of third parties linked to participant THIRD PARTIES	(L) Max.EU Contribution / € THIRD PARTIES	(M) Total Costs for BENEFICIARY & THIRD PARTIES (=H+K)	(N) Max.EU Contribution / € BENEFICIARY & THIRD PARTIES (=J+L)	(O) Requested EU Contribution / € BENEFICIARY & THIRD PARTIES
1	Bsc	ES	971300	148334	0	0	0	279908,50	0	1399542,50	100	1399542,50	0	0	1399542,50	1399542,50	699771,25
2	Cimne	ES	378000	23100	0	0	0	100275,00	0	501375,00	100	501375,00	0	0	501375,00	501375,00	250687,50
3	Fzj	DE	720000	23100	0	0	0	185775,00	0	928875,00	100	928875,00	0	0	928875,00	928875,00	464437,50
4	Upv	ES	324000	23100	0	0	0	86775,00	0	433875,00	100	433875,00	0	0	433875,00	433875,00	216937,50
5	Bull	FR	629004	50100	0	0	0	169776,00	0	848880,00	100	848880,00	0	0	848880,00	848880,00	297108,00

6	Dtok Lab S.r.l.	IT	311040	16800	0	0	0	81960,00	0	409800,00	100	409800,00	0	0	409800,00	409800,00	143430,00
7	Fondazione Cmcc	IT	504000	28100	0	0	0	133025,00	0	665125,00	100	665125,00	0	0	665125,00	665125,00	332562,50
8	Inria	FR	225000	16800	0	0	0	60450,00	0	302250,00	100	302250,00	0	0	302250,00	302250,00	151125,00
9	Sissa	IT	225000	16800	0	0	0	60450,00	0	302250,00	100	302250,00	0	0	302250,00	302250,00	151125,00
10	Psnc	PL	72000	16800	0	0	0	22200,00	0	111000,00	100	111000,00	0	0	111000,00	111000,00	55500,00
11	Uma	ES	159800	16800	0	0	0	44150,00	0	220750,00	100	220750,00	0	0	220750,00	220750,00	110375,00
12	Istituto Nazionale Di Geofisica E Vulcanologia	IT	207360	16800	0	0	0	56040,00	0	280200,00	100	280200,00	0	0	280200,00	280200,00	140100,00
13	Awi	DE	295200	21050	0	0	0	79062,50	0	395312,50	100	395312,50	0	0	395312,50	395312,50	197656,25
14	Eth Zürich	CH	150385	16800	0	0	0	41796,25	0	208981,25	100	208981,25	0	0	208981,25	208981,25	104490,63
15	Siemens Aktiengesellschaft	DE	336000	16800	0	0	0	88200,00	0	441000,00	100	441000,00	0	0	441000,00	441000,00	154350,00

16	Ngi	NO	154400	16800	0	0	0	42800,00	0	214000,00	100	214000,00	0	0	214000,00	214000,00	107000,00
	Total		5662489	468084	0	0	0	1532643,25	0	7663216,25		7663216,25	0,00	0,00	7663216,25	7663216,25	3576656,13

4 - Ethics

1. HUMAN EMBRYOS/FOETUSES		Page
Does your research involve Human Embryonic Stem Cells (hESCs)?	○ Yes ◉ No	
Does your research involve the use of human embryos?	○ Yes ◉ No	
Does your research involve the use of human foetal tissues / cells?	○ Yes ◉ No	
2. HUMANS		**Page**
Does your research involve human participants?	○ Yes ◉ No	
Does your research involve physical interventions on the study participants?	○ Yes ◉ No	
3. HUMAN CELLS / TISSUES		**Page**
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	○ Yes ◉ No	
4. PERSONAL DATA		**Page**
Does your research involve personal data collection and/or processing?	○ Yes ◉ No	
Does your research involve further processing of previously collected personal data (secondary use)?	○ Yes ◉ No	
5. ANIMALS		**Page**
Does your research involve animals?	○ Yes ◉ No	
6. THIRD COUNTRIES		**Page**
In case non-EU countries are involved, do the research related activities undertaken in these countries raise potential ethics issues?	○ Yes ◉ No	
Do you plan to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)?	○ Yes ◉ No	
Do you plan to import any material - including personal data - from non-EU countries into the EU?	○ Yes ◉ No	
Do you plan to export any material - including personal data - from the EU to non-EU countries?	○ Yes ◉ No	
In case your research involves low and/or lower middle income countries, are any benefits-sharing actions planned?	○ Yes ◉ No	
Could the situation in the country put the individuals taking part in the research at risk?	○ Yes ◉ No	

7. ENVIRONMENT & HEALTH and SAFETY		Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants?	○ Yes ◉ No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	○ Yes ◉ No	
Does your research involve the use of elements that may cause harm to humans, including research staff?	○ Yes ◉ No	
8. DUAL USE		**Page**
Does your research involve dual-use items in the sense of Regulation 428/2009, or other items for which an authorisation is required?	○ Yes ◉ No	
9. EXCLUSIVE FOCUS ON CIVIL APPLICATIONS		**Page**
Could your research raise concerns regarding the exclusive focus on civil applications?	○ Yes ◉ No	
10. MISUSE		**Page**
Does your research have the potential for misuse of research results?	○ Yes ◉ No	
11. OTHER ETHICS ISSUES		**Page**
Are there any other ethics issues that should be taken into consideration? Please specify	○ Yes ◉ No	

I confirm that I have taken into account all ethics issues described above and that, if any ethics issues apply, I will complete the ethics self-assessment and attach the required documents. ☒

How to Complete your Ethics Self-Assessment

5 - Call-specific questions

Extended Open Research Data Pilot in Horizon 2020

If selected, applicants will by default participate in the Pilot on Open Research Data in Horizon 2020[1] , which aims to improve and maximise access to and re-use of research data generated by actions.

However, participation in the Pilot is flexible in the sense that it does not mean that all research data needs to be open. After the action has started, participants will formulate a Data Management Plan (DMP), which should address the relevant aspects of making data FAIR – findable, accessible, interoperable and re-usable, including what data the project will generate, whether and how it will be made accessible for verification and re-use, and how it will be curated and preserved. Through this DMP projects can define certain datasets to remain closed according to the principle "as open as possible, as closed as necessary". A Data Management Plan does not have to be submitted at the proposal stage.

Furthermore, applicants also have the possibility to opt out of this Pilot completely at any stage (before or after the grant signature). In this case, applicants must indicate a reason for this choice (see options below).

Please note that participation in this Pilot does not constitute part of the evaluation process. Proposals will not be penalised for opting out.

We wish to opt out of the Pilot on Open Research Data in Horizon 2020.	○ Yes	◉ No

Further guidance on open access and research data management is available on the Funding & Tenders portal: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm and in general annex L of the Work Programme.

[1] *According to article 43.2 of Regulation (EU) No 1290/2013 of the European Parliament and of the Council, of 11 December 2013, laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" and repealing Regulation (EC) No 1906/2006.*

Enabling Dynamic and Intelligent Workflows in the Future EuroHPC Ecosystem

Call topic: EuroHPC-02-2019: HPC and data-centric environments and application platforms

List of participants

Participant No	Participant organisation name	Short Name	Country
P1 (Co)	Barcelona Supercomputing Center	BSC	Spain
P2	International Centre for Numerical Methods in Engineering	CIMNE	Spain
P3	Forschungszentrum Jülich GmbH	FZJ	Germany
P4	Universitat Politècnica de València	UPV	Spain
P5	Bull SAS	ATOS	France
P6	DtoK Lab S.r.l.	DtoK	Italy
P7	Fondazione Centro Euro-Mediterraneo Sui Cambiamenti Climatici	CMCC	Italy
P8	French Institute for Research in Computer Science and Automation	INRIA	France
P9	Scuola Internazionale Superiore di Studi Avanzati	SISSA	Italy
P10	Poznan Supercomputing and Networking Center	PSNC	Poland
P11	Universidad de Málaga	UMA	Spain
P12	Istituto Nazionale di Geofisica e Vulcanologia	INGV	Italy
P13	Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung	AWI	Germany
P14	Eidgenössische Technische Hochschule Zürich	ETHZ	Swiss
P15	Siemens AG	Siemens	Germany
P16	Stiftelsen Norges Geotekniske Institutt	NGI	Norway

Table of Contents

Section 1. Excellence

1.1 Objectives

The goals of a large number of complex scientific and industrial applications are deeply linked to the **effective use of high-performance computing infrastructures** and the efficient **extraction of knowledge** from vast amounts of data. This involves large amounts of data coming from different sources that follow a process composed of pre-processing steps for intermediate data curation and preparation for subsequent computing steps, and later analysis and analytics steps applied to produce results. However, workflows are currently fragmented in multiple components, using different programming models, with different processes for computing and data management.

While a large number of current application workflows follow a brute force approach, where the design or research space is blindly explored, new intelligent approaches that make the best use of the computational resources are urgently needed. In this sense, the **carbon footprint** of ICT processes is a concern, and its reduction is one of the objectives of the Horizon Europe framework program.

The project aims to demonstrate through three application Pillars with high **industrial and social relevance**: **manufacturing, climate, and urgent computing for natural hazards**; how the realization of forthcoming **efficient HPC and data-centric applications** can be developed with new **workflow technologies**.

In this sense, eFlows4HPC aims to propose **workflow technologies and services** that enable the integration of **HPC simulation and modelling with big data analytics and machine learning** and to **demonstrate this through the different Pillars' application areas**. The project will integrate existing workflow interfaces, programming models, machine learning, and data analytics libraries to provide a uniform, easy to use platform that enables the exploitation of future large-scale systems. The type of workflows that we aim to support is of a very **dynamic nature**, both because the workflows will be instantiated at execution time and because it is envisaged that the workflows will evolve in an **intelligent manner** in response to results obtained in the ongoing execution. This could involve changes in configuration or in enacting different sets of computations to those initially foreseen, based upon information resulting from earlier steps in the process. For example, when exploring a solution area, a set of initial simulations are conducted and, using data analytics, some options are discarded, while others are explored in more detail. In conclusion, **the solution area is explored intelligently**, making **efficient** usage of the resources, **both in time and energy**, and enabling a faster time-to-solution.

The following characteristics are foreseen in the workflows platform: support for the integration of HPC simulation and modelling, data analytics and machine learning; support for dynamic workflows that can change their behaviour during the execution; support for dynamic resource management depending on the actual workload needs; support for data-streaming; support for persistent storage beyond traditional file systems. The type of computing platforms that we are considering will be centred on large **HPC systems** (PRACE tier-0 and tier-1 and EuroHPC pre-exascale

systems) but will also be connected to **external devices or instruments** for data acquisition and including **cloud-based solutions,** like a **data logistics service** for the on-demand, self-service, automatic movement and pre-process data. The project will also devote efforts to consider the **heterogeneity** of the systems and how specialized architectures can be used in each of the Pillars' use cases. A special case will be the study of the requirements of the Pillars' use cases in the context of the new **European Processor Initiative (EPI)**, by optimizing specific kernels or parts of applications to its architecture.
The project also aims at proposing and developing the **HPC Workflows as a Service (HPCWaaS)** concept, as a means for **widening the access to HPC** from user communities. The goal is to provide methodologies and tools that enable to share and reuse existing workflows and that assist when adapting workflow templates to create new workflow instances.
The outcomes of the project will include an open **eFlows4HPC platform**, composed of the **eFlows4HPC software stack** in the form of interoperable components and associated services, and of the **HPCWaaS methodology**; a **workflow registry and model repository** and **a set of (re)usable workflows derived from the project Pillars' use cases**. The required and available data assets will be also clearly described. The project will involve application stakeholders (as partners) in the pillar use cases, with the goal of defining complex workflows that include modelling and simulation together with High-Performance Data Analytics (HPDA) and Artificial Intelligence (AI). The final goal is the **adoption** of the solutions defined in the project by the **user communities**, both by enabling **impact on industrial cases and their exploitation in future HPC systems**, by means of providing secure and simple access to these solutions, in some cases in the form of community services. To reinforce this, the **Centres of Excellence (CoEs) ChEESE, ESiWACE and EXCELLERAT** will be involved in the process, and in coordination with the **Focus CoE** for transfer activities to other communities.

The eFlows4HPC proposal is unique in its nature and has the potential to become strategic for the EuroHPC purposes.

Main Project Objective: The project aims to deliver a workflow platform that consists of the eFlows4HPC software stack and an additional set of services that will enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The platform will also include methodologies to widen access to HPC by the different communities through the HPCWaaS concept. The project will demonstrate through three application Pillars with high industrial and social relevance: manufacturing, climate and urgent computing for natural hazards, how the realization of forthcoming efficient HPC and data-centric applications can be developed with new workflow technologies.

Project's specific objectives

The project's main objective is decomposed in multiple specific objectives classified into:

- Scientific and Technological objectives (STO) focused on the delivery of a workflows software stack and added value services, to be used as the basis for exploitation by the HPC centres in Europe.
- Pillar-specific Scientific objectives (PSO) focused on the delivery of application workflows and workflow templates, that can be exploited by the current application stakeholders involved in the project and by the corresponding communities in its usage of HPC.
- Societal and Industrial objectives (SIO) focused on the pre-commercial evaluation and validation of the project solutions and on the exploitation by communities of the project results.

	Scientific and Technological objectives (STO)
STO1	To deliver an open European workflows software stack (eFlows4HPC) for the development and management of complex workflows, by integrating and extending existing European components and enabling their energy-efficient execution.
STO2	To enable the definition of complex workflows integrating HPC simulation and modelling with high-performance data analytics and machine learning in scientific and industrial applications.
STO3	To define an open methodology that widens and eases the use of workflows by existing and new HPC communities and users — HPC Workflows as a Service (HPCWaaS). The solution will adhere to the necessary standards for security and authorization required in HPC environments.

STO4	To provide means to increase openness, transparency, reusability, and reproducibility of computation results by formalizing data inputs and storing created models along with a proper metadata description.
STO5	To optimize specific application kernels from the Pillars' use cases to efficiently accommodate the EPI and other emerging platforms and architectures (GPUs, FPGAs, and custom accelerators), addressing energy efficiency goals as well.

	Pillar-specific Scientific objectives (PSO)
PSO1	To develop area-specific workflows following a multidisciplinary approach where experts from different areas are involved.
PSO2	To enable interaction between simulation and AI by constructing simulation-based Digital Twins for manufactured products and use ML approaches to enable fast yet deterministic response.
PSO3	To develop and optimize specific and complex climate simulation workflows, enabling dynamic and adaptive ensemble climate runs.
PSO4	To reduce workflow-level bottlenecks in urgent simulations for natural hazards and climate, and reach technological maturity for their implementation as a service.
PSO5	To develop Earth System Model workflows transparently, seamlessly and efficiently integrating numerical models, HPDA and ML software components.
PSO6	To enable the smart integration (with respect to time-to-solution and energy consumption metrics) of data-intensive versus data-driven approaches for feature extraction (e.g., tropical cyclone tracking), into climate simulation workflows.

	Societal and industrial objectives (SIO)
SIO1	To introduce fast, deterministic Reduced-Order Digital Twins in the industrial pipeline. To employ HPC in the training phase with the aim of ultimately delivering computationally inexpensive models enabling faster-than-real-time predictions.
SIO2	To propose strategies to reduce the time to evaluate the impact of a natural hazard.
SIO3	To promote and exploit European technology in the Pillars' use cases.
SIO4	To establish a mechanism to feedback project results, in particular, to target Pillars' communities through the corresponding CoEs, and for a wider community through the CSA Focus CoE.

Means of verification

At the **end of the project**, the following **means of verification** will validate the aforementioned objectives.

	Means of verification (MoV)
MoV1	eFlows4HPC workflow platform (eFlows4HPC software stack, HPCWaaS and associated services) deployed in at least 2 European infrastructures
MoV2	A registry of workflows and workflow templates available to existing and new HPC communities and the corresponding means to manage and deploy them through the HPCWaaS methodology.
MoV3	A repository of AI trained models and a catalogue of libraries and services that can be used by new and existing communities as components of new workflows.

MoV4	A set of application kernels from manufacturing, climate, and natural hazards optimized for the EPI and other emerging platforms and architectures (GPUs, FPGAs, and custom accelerators).
MoV5	An open report on implementation strategies and recommended policies of urgent computing within the European HPC ecosystem.

1.2 Relation to the work programme

The following table reflects how the proposal is contributing to the different call aspects and which of the project objectives address this contribution.

eFlows4HPC contributions to call EuroHPC-02-2019
Call extract: *to support the development of High Performance Computing (HPC) and data driven HPC software environments and application-oriented platforms to generate innovation and value creation in sectors of societal and industrial relevance for Europe.*
eFlows4HPC will contribute with workflow methodologies integrating HPC, HPDA and AI; and with sample applications from manufacturing, climate, and urgent computing. The project aims at establishing standard methodologies that make easier the usage and development of applications in those areas and extend the use of HPC in Europe. **Addressed objectives:** Main objective
Call extract: *Proposals should address the development of energy-efficient HPC solutions.*
The project will be based on the eFlows4HPC software stack, which will provide a technology that enables the development of very dynamic workflows that can change according to the partial results of the application. The stack will leverage on dynamic runtimes able to obtain energy-efficient application solutions. The project will also deliver repositories, registries and catalogues that will facilitate the reuse of computation results and thus reduce the energy wasted in redoing the work. Heterogeneous components will be targeted in the project to enable further hardware-level energy efficiency. **Addressed objectives:** STO1, STO5, PSO6
Call extract: *supporting the adoption of applications with industrial and societal relevance for Europe on evolving HPC hardware and system software/programming environments. The use of HPC solutions to generate innovation and value creation should be clearly demonstrated (for instance in manufacturing, farming, health, mobility, natural hazards, climate, energy, space, finance or cybersecurity).*
The project will develop workflow methodologies and solutions for manufacturing, climate and natural hazards targeting PRACE Tier-0 and Tier-1 systems, and considering new architectures to come (EPI, Fugaku architecture). The results of the Pillars' developments will be fed to the relevant communities, for example through the ChEESE, EXCELLERAT and ESiWACE CoEs, and to the industry involved in the project. Beyond the mentioned application areas, the workflow stack and services delivered by eFlows4HPC aim to be transversal to the application domain, and therefore reusable by other communities. **Addressed objectives:** PSO2, PSO3, PSO4, PSO5, PSO6, SIO1, SIO2, SIO3
Call extract: *providing secure and simple access and service provisioning to relevant stakeholders based on such HPC solutions.*
The proposal aims to provide a methodology that widens and eases the use of workflows by existing and new HPC communities - HPC Workflows as a Service (HPCWaaS). The solution will adhere to the necessary standards for security and authorization required in HPC environments. The project will also put in place registries, repositories and catalogues that will simplify the development and deployment of new applications. **Addressed objectives:** STO3, STO4, SIO4
Call extract: *These developments should be driven by complex application workflows, for instance High Performance Data Analytics (HPDA), combining artificial intelligence and simulation modelling, exploiting underlying hardware heterogeneity/modularity, integrating cloud-based solutions etc., and should offer solutions to key application areas including industrial use cases.*

The project objectives will provide a software stack and services with interfaces and environments that enable the definition and execution of complex workflows comprising HPC, HPDA and AI. eFlows4HPC will be demonstrated in the project with use cases from 3 Pillars in key areas, which include simulation models and with industrial and societal impact: manufacturing, climate and natural hazards. This will be achieved by leveraging the two partner-provided tools: the TOSCA-based[i] Ystia orchestrator combined with the PyCOMPSs workflow environment. Also, heterogeneous hardware will be targeted in the project with the support and adaptation of European libraries such as EDDL and ECV[ii]. **Addressed objectives:** Main objective, STO1, STO2, PSO2, PSO3, PSO4, PSO5, PSO6
Call extract: *These developments may also promote the efficient use of platforms and architectures best suited for the targeted use cases and applications (e.g. accelerated platforms).*
Specific application kernels from the Pillars' use cases will be optimized for the efficient use of the EPI and other specific platforms (GPUs, FPGAs, and custom accelerators). Models often require specific platforms to speed-up the training phase, but once created can be used to make predictions and classifications on regular hardware. This will be enabled by the model repository with deployment option. **Addressed objectives:** STO4, STO5, SIO4
Call extract: *The required and available data assets should be clearly described.*
The data used in the Pillars' use cases is related to the representation of mathematical models of the objects to be simulated (i.e., finite element meshes); high-resolution climate simulations; and parametric and waveform data for earthquakes and tsunamis. The Pillars' data assets are described in section 1.3.5. **Addressed objectives:** STO4
Call extract: *Actions should include co-design in close cooperation with the scientific disciplines to explore and demonstrate the technical feasibility and value of advanced workflows, e.g. mixed/integrated simulation, HPDA & AI,...*
The project methodology will be based on co-design between different scientific areas with experts from the areas of HPC, artificial intelligence, data analytics, workflow environments, data management, computational sciences, computational mechanics, and others, cooperating all together towards the project solutions (see section 3.1). **Addressed objectives:** main objective, PSO1
Call extract: *ensuring wide adoption in production use.*
The project will design the methodology and develop the corresponding tools to enable the delivery of HPC Workflows as a Service (HPCWaaS) that aim to widen the adoption of HPC from communities (STO3). To promote this adoption, the project will involve in its activities the Centres of Excellence (especially ChEESE, ESiWACE, and EXCELLERAT), and coordinate with Focus CoE for these activities and the industrial community stakeholders. **Addressed objectives:** STO3, STO4, SIO4

1.3 Concept and Methodology

1.3.1 Overall Approach

The scientific process has been described as composed by three inference steps[1]: **abduction** (i.e., guessing at an explanation), **deduction** (i.e., determining the necessary consequences of a set of propositions), and **induction** (i.e., making a sampling-based generalisation). These key logical elements have been presented in the recently published white paper[2] by the Big-Data and Extreme-Scale Computing (BDEC)[iii], an international initiative participated by the proposal PI as one of the European executive members, which focuses on the convergence of HPDA and HPC. While the abduction and induction phases imply the use of analysis and analytics processes (HPDA techniques), the deduction phase is typically an HPC process. However, the three different steps of the scientific process have been realised until now with separated methodologies and tools, with a lack of integration and lack of common view of the whole process. The main BDEC recommendation is to address the basic problem of the split between the two

[i] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=tosca
[ii] https://www.deephealth-project.eu
[iii] https://www.exascale.org/bdec/

paradigms: the HPC/HPDA software ecosystem split. In addition, current international roadmaps, including the BDEC, focus on combining HPC with AI, itself tightly linked to the big data revolution.
Another observation is that the usage of HPC resources by scientific workflows is often done in a brute force manner where a large number of simulations or modelling jobs are submitted, generating themselves a large amount of data which are later analysed/processed in a decoupled process. There is a need for smarter workflow approaches, able to use HPC in a more **energy-efficient** way but also able to perform the different HPC/HPDA/AI steps in a more integrated form.
The situation is kind of similar in the context of **industrial applications**: for example, for the Pillar I use case, using current technologies "Full Order Models" (FOM) are generated for increasingly complex designs, generating a large amount of data that is processed in later steps to generate Reduced Order Models (ROM) that can be used in the construction of digital twins. A more integrated approach will streamline the solution of FOM problems opening the door to adaptive algorithms. This, in turn, will allow faster and more reliable ROM, thus improving the impact over industry.
eFlows4HPC aims at designing and implementing a **European workflow platform** that enables the design of complex applications that integrate HPC processes, data analytics and artificial intelligence, making use of the HPC resources in an easy, efficient and responsible way as well as enabling the accessibility and reusability of applications to reduce the time to solution. Figure 1 shows the overall approach to achieve the main project goal. As the main outcome, the project will deliver the **eFlows4HPC software stack** which integrates different components to provide an overall workflow management system. On top of this software stack, we will build an **HPC Workflow as a Service** (HPCWaaS) platform to facilitate the reusability of these complex workflows in federated HPC infrastructures[iv]. The HPCWaaS platform and the eFlows4HPC software stack will be validated by **use cases** organised in three Pillars which represent the main sectors we target in the project. These Pillars will develop methodologies and complex workflows that will be used as pilots for the **industry, climate and natural hazards communities** represented by different Centres of Excellence.

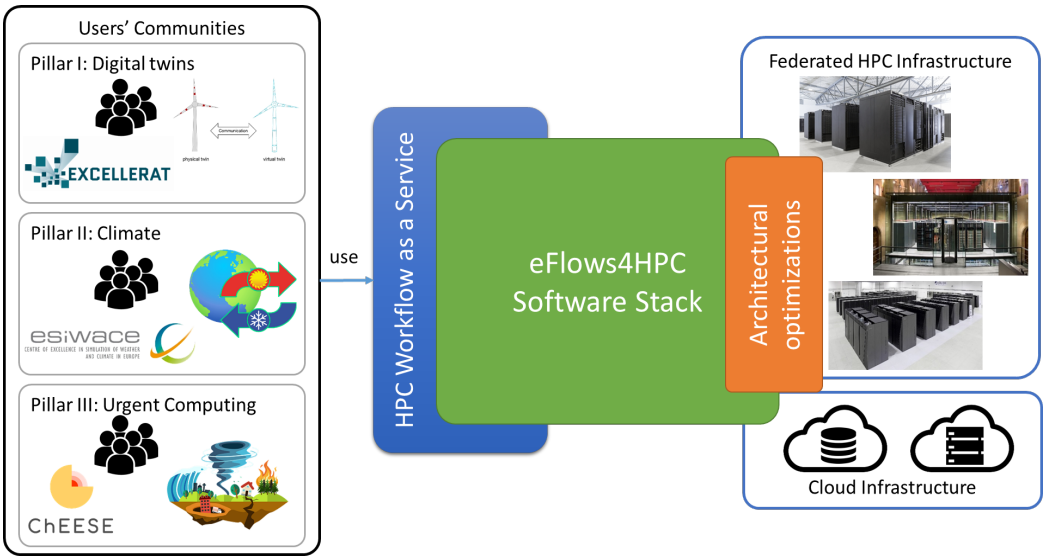


Figure 1: eFlows4HPC project overall approach

1.3.2 Methodology

Next we present the activities foreseen in the project to achieve the previously proposed innovation.

eFlows4HPC software stack
The eFlows4HPC software stack will be composed of the integration of existing software components which are organised in different layers (Figure 2). The first layer consists of a set of repositories, catalogues and registries to facilitate the accessibility and re-usability of the implemented workflows and their core software components (HPC libraries, DA/ML frameworks) and ML models. The second layer provides the syntax to implement these complex workflows combining typical HPC simulations with HPDA and ML, and a description about how the workflows are deployed in the computing infrastructure. Finally, the lowest layers provide the functionalities to orchestrate the deployment of the workflow components and controlling the execution in federated computing infrastructures.

[iv] In this proposal, a federation refers to a set of HPC resources geographically distributed used in collaboration for a workflow execution.

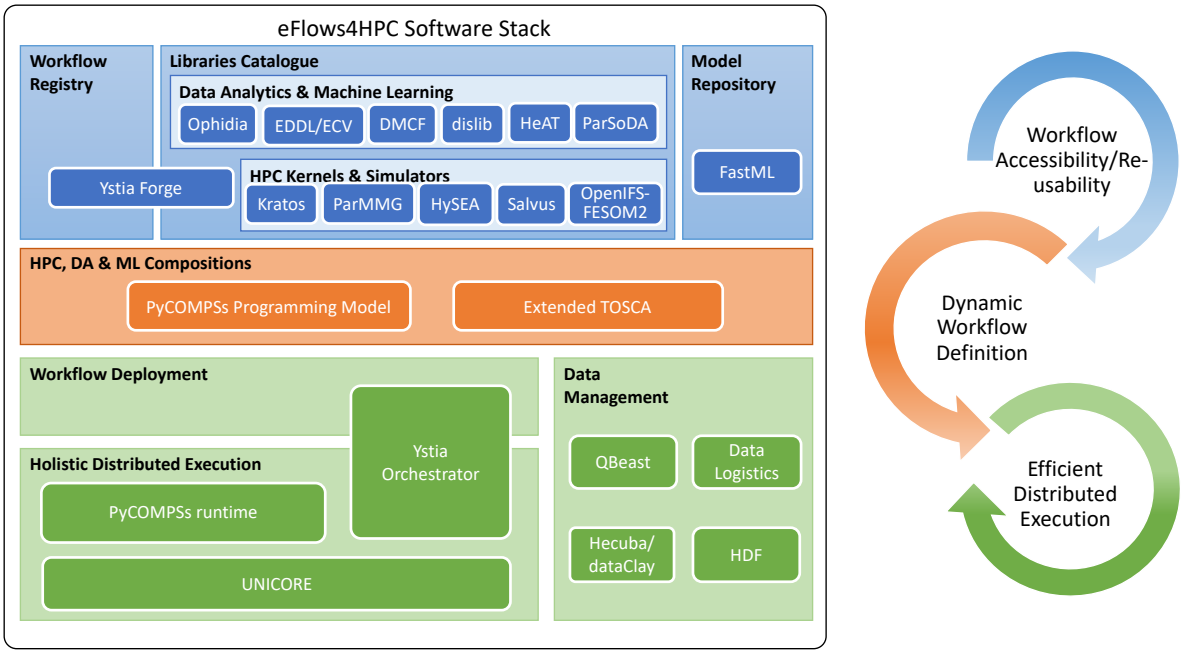


Figure 2: eFlows4HPC Software Stack

One of the core functionalities of the software stack is the definition of the complex workflows that combine HPC, HPDA and ML frameworks. In order to achieve this functionality, we propose a combination of an Extended TOSCA syntax and the PyCOMPSs programming model.

On the one hand, TOSCA (an orchestration standard) provides a way to describe the overall workflow components, how they are deployed, configured, run, and linked to each other. It allows managing the global application lifecycle. TOSCA allows developers to specify which software and services are required; how each component of a workflow should be created, configured, started, stopped, and deleted; which infrastructure components they should be deployed on; and what could be performed in parallel. TOSCA provides deployment workflows that can be customised. The intent is to leverage this and potentially extend it to support dynamic workflows, failover workflows, and data-aware workflows.

On the other hand, the PyCOMPSs programming model will provide the logic of the different components of the overall workflow. PyCOMPSs is a task-based programming model that enables the development of workflows that can be executed in parallel on distributed computing platforms as easily as programming sequential Python scripts, offering the programmer the illusion of a single shared memory and storage space. While the PyCOMPSs task orchestration code needs to be written in Python, it supports different types of tasks, such as Python methods, external binaries, multi-threaded (internally parallelised with alternative programming models such as OpenMP or pthreads), or multi-node (MPI applications). Thanks to the use of Python as programming language, PyCOMPSs naturally integrates well with data analytics and machine learning libraries, most of them offering a Python interface. PyCOMPSs also supports reading/writing streamed data.

Once the complex workflow is defined as explained above, the eFlows4HPC software stack aims at integrating different existing tools to provide a holistic workflow deployment and execution in federated computing HPC infrastructures. This integration will be performed by several components cooperating at different levels. At the highest level, the Ystia Orchestrator (Yorc) is in charge of orchestrating the deployment of the main workflow components in the computing infrastructures and managing their lifecycle (starting services and invoking management actions, submitting jobs at the HPC infrastructure...) as described in the TOSCA description.

At a lower level, the COMPSs runtime will manage the execution of the workflow components implemented with the PyCOMPSs programming model. At runtime, it generates a task-dependency graph by analysing the existing data dependencies between the tasks defined in the Python code. The task-graph encodes the existing parallelism of the workflow, which is then scheduled and executed by the COMPSs runtime in the resources already deployed by Yorc. The COMPSs runtime is also able to react to tasks failures and to exceptions in order to adapt the behaviour accordingly. These functionalities, together with similar features provided by Yorc at a higher level, offer the possibility of designing a new category of workflows with very dynamic behaviour, that can change their configuration at execution time upon the occurrence of given events, such as failures or exceptions. At the lowest level, UNICORE will be in charge of managing the federation of HPC compute and data resources in order to make them available to users in a secure way. Yorc and the COMPSs runtime will be integrated with UNICORE in order to automate the deployment and execution of the implemented workflows in these infrastructures.

Regarding data management, the eFlows4HPC stack provides a set of components to manage and simplify the integration of large volumes of data from different data sources and locations with the workflow execution. First, the stack provides the Data Logistic Service, a tool to define and execute data pipelines to collect and integrate distributed data before they are used for processing. Second, it also provides two solutions for persistent storage: Hecuba (based on key-value databases) and dataClay (object-oriented distributed storage). These solutions can be used in PyCOMPSs applications to store application objects as persisted objects in new memory devices such as NVRAM or SSDs, enabling to keep the data after the execution of the application. This changes the paradigm of persistent storage in HPC, dominated by the file system, to other, more flexible approaches. By using persisted objects, application patterns such as producer-consumer, in-situ visualisation or analytics, can be easily implemented. Finally, the data management part of the stack is completed with QBeast, a scalable multidimensional indexing system and sampling mechanism for NoSQL databases, and the Helmholtz Data Federation (HDF) which manages the federated access to storage resources.

HPC Workflows as a Service (HPCWaaS)

One of the main current barriers to the adoption of HPC is the complexity of deploying and executing the workflows in federated HPC environments. Usually, users are required to perform software installations in complex systems which are beyond their technical skills. Therefore, having the workflows ready for execution in a supercomputer could take some time and human resources. If it has to be replicated for several clusters, the required time and resources will increase. To widen the access to HPC for newcomers, and, in general, to simplify the deployment and execution of complex workflows in HPC systems, eFlows4HPC will design and develop a mechanism to deploy HPC Workflows as a Service (HPCWaaS) following a similar concept as the Function as a Service (FaaS) has in the Cloud, but applying it for workflows in federated HPC environments. The goal is to provide methodologies and tools that enable to share and reuse existing workflows and that assist when adapting workflow templates to create new workflow instances. eFlows4HPC will leverage the Ystia orchestrator based on the TOSCA standard to develop this methodology and environment.

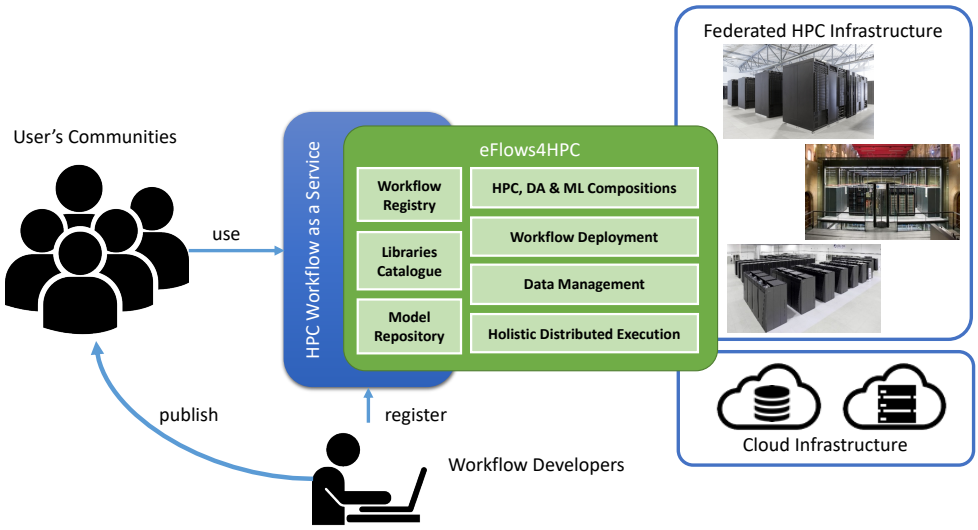


Figure 3: Architecture of HPC Workflow as a Service (HPCWaaS)

Figure 3 shows the architecture that is devised, where the HPC Workflow as a Service offering is built on top of the eFlows4HPC software stack in order to provide the required functionality as a Service. The usage of the HPCWaaS will have two ways of interaction, one for workflow developers and another for workflow user communities. Details about the proposed interactions are shown in Figure 4. In the development phase, the service will offer a catalogue for registering the libraries and a model repository, providing workflow developers with a means to store the HPC, HPDA and ML libraries and models in a way so that they can be easily reused and deployed in HPC infrastructures. Workflow developers can define the complex workflows using the PyCOMPSs programming model and TOSCA descriptions. Once a workflow has been defined, the developer can register and publish it in the registry to make it available to the whole community of the target sector. Other users can later reuse it, which will be easily deployed and executed using the eFlows4HPC stack and services. Additionally to instances of specific workflows, it will be possible to register workflow templates. All the process will be subject to the security authorisation and authentication rules required in such infrastructures. All necessary data pipelines, which collect and integrate distributed data before being processed, will be performed by the Data Logistics Service. While the methodologies that we aim to design will, in most cases, be usable in cloud environments, the proposal aims at mainly addressing HPC infrastructures for computing. The Cloud will be considered for data storage.

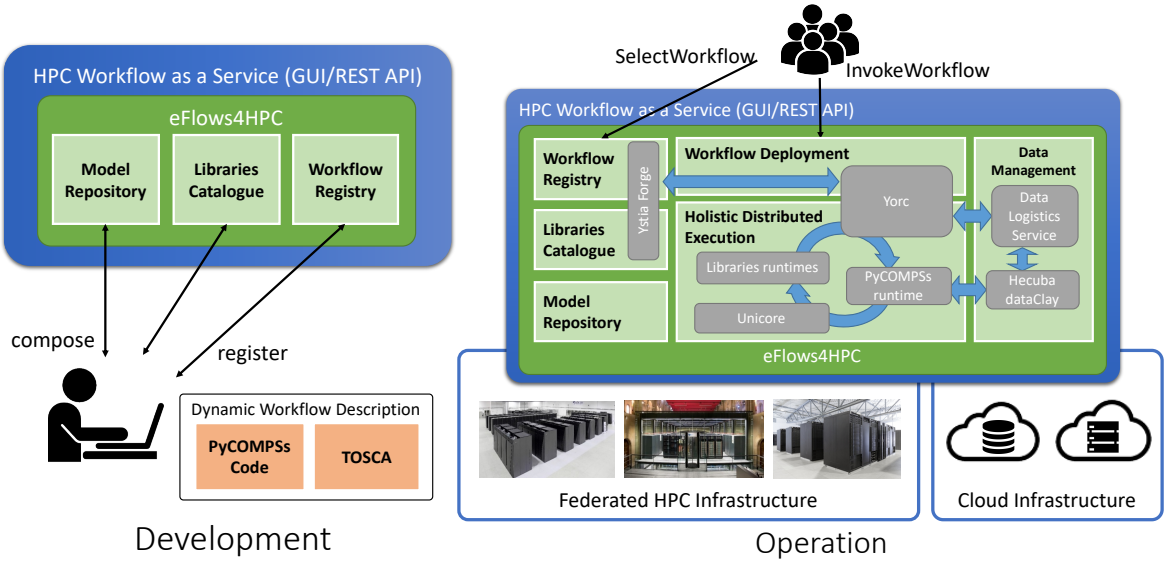


Figure 4: HPCWaaS usage overview in Development (left side) and Operation (right side) phases

Deployment in HPC systems and optimization for heterogeneous infrastructures available for eFlows4HPC

eFlows4HPC will benefit from an ecosystem of infrastructures and initiatives that exist in the partners ecosystem. First, the partners will have access to pre-exascale EuroHPC JU systems, such as MareNostrum 5 that will be deployed in October 2020 at the BSC premises[v]. The partners will also have access to the JUWELS modular Tier-0/1 Supercomputer at FZJ, to a 1.7 PFlops HPC system at PSNC for the Pillar III tests, and to the Athena and Zeus supercomputers at CMCC. UPV will use and enable partners access to the MANGO prototype made of 96 interconnected high-end FPGA boards, which will be used to deploy energy-efficient optimizations of target use cases within the project. This prototype has been manufactured and built from the H2020 MANGO project with the aim to explore future energy-efficient heterogeneous HPC architectures. In addition, being BSC one of European Processor Initiative (EPI) partners, will have access to the MareNostrum Exascale Emulation Platform (MEEP), which will develop an FPGA-based system that will emulate the EPI and to a real platform in a Fujitsu cluster (Fugaku architecture) that will be installed at BSC in 2020. This cluster will be based on the ARM processor which has a very similar architecture to the EPI. The HDF Cloud – Helmholtz Data Federation Cloud Resources from FZJ will be available for data storage and management, a data infrastructure from PSNC (42 PB) and a Tier-2 storage system (4PB) at CMCC.

In this sense, another activity performed in the project will focus on the optimization of application kernels of the HPC, HPDA and ML frameworks available in the catalogue and used by the different Pillars for specific aspects of future HPC architectures, following co-design methodologies. The optimizations will consider heterogeneous architectures, including GPUs, FPGAs and custom accelerators, with special emphasis on the EPI and the usage of novel storage solutions (i.e. Intel Optane DC).

Validation with Pillars' use cases

eFlows4HPC consortia has selected for demonstrating the workflow platform use cases for three thematic Pillars on manufacturing (Pillar I), climate (Pillar II) and natural hazards (Pillar III). Pillar I will focus on the construction of **Digital Twins for the prototyping of complex manufactured objects** integrating state-of-the-art adaptive solvers with machine learning and data-mining, contributing to the Industry 4.0 vision. Pillar II will develop innovative **adaptive workflows for climate modelling** able to make efficient use of computing resources by means of performing a dynamic pruning of the simulations thanks to the dynamic features of PyCOMPSs. It will also leverage the workflow software stack to study Tropical Cyclone (TC) track exploiting multi-model analysis in the context of the CMIP6 experiment, and in-situ analytics. Pillar III will explore the **modelling of natural catastrophes**, in particular earthquakes and their associated tsunamis, shortly after such an event is recorded. Leveraging two existing workflows, the Pillar will work on integrating them with the eFlows4HPC software stack and on producing policies for **urgent access to supercomputers**.

[v] The access to MareNostrum is granted to BSC; additional allocation for the project will be requested through peer-review. Similar situation applies to JUWELS.

1.3.3 Baseline technologies and Technology Readiness Level (TRL)

As an innovation action project, eFlows4HPC is expected to integrate into the final solution a plethora of methods, libraries and tools coming from **open source** initiatives, past and ongoing projects, as well as other services that are highly operational. In this context, besides its own innovations, eFlows4HPC will reuse existing artifacts and will extend them if necessary, to accelerate development, eliminate risks and provide users with service environments they are already familiar with. The overall TRLs of the different project components, even from the very beginning, is quite high and the viability and maturity of tools and components, which are both of high TRL and MRL (Market Readiness Levels), guarantee that the final results will be brought up to the market or in production in EuroHPC sites quite fast and in a high-fidelity operational mode. The basic components for the proposed eFlows4HPC workflow stack, as shown in the table below, are already of pre-commercialisation levels (TRL 6-8), which is a key factor for exploitation of the final project's results. For that reason, the aim of the project is to provide the **integrated eFlows4HPC software stack at TRL 7** (system prototype demonstration in an operational environment) and in the case of the **HPC Workflow as a Service offering with a minimum TRL 6** (technology demonstrated in a relevant environment).

Software Stack baseline components

Technology	Short description	Current TRL
PyCOMPSs/ COMPSs	Task-based programming environment that enables the development of applications and workflows that are executed in parallel in distributed computing platforms.	7
Ystia/Yorc	TOSCA based orchestrator to manage applications lifecycle on hybrid infrastructures (IaaS, CaaS, HPC clusters)	8
Ystia/Forge	Catalogue of TOSCA components and applications	8
dataClay	Distributed object store. Avoids data transformations and movements from the store to the application space. Integrated with PyCOMPSs.	7
Hecuba	Set of tools and interfaces that aims to facilitate programmers an efficient utilization of key-value data stores	6
Qbeast	Highly scalable multidimensional indexing system that provides NoSQL databases with an efficient indexing and sampling mechanism	6
UNICORE/ PyUNICORE	UNICORE is a federation software suite that makes HPC compute and data resources accessible for end users and applications in a secure way. PyUNICORE is a Python library wrapping the UNICORE RESTful APIs	9/7
Helmholtz Data Federation (HDF)	HDF is a nation-wide cloud environment enabling federated access to compute and storage resources.	9
Data Logistics Service	Data Logistic Service is a tool to define, schedule, and execute in a reproducible way data pipelines. It collects and integrates distributed data before they are used for processing.	5
Model Repository	Model Repository enables sharing of computation artifacts (e.g. trained models) with additional metadata describing model parameters and evaluation metrics.	4
FastML Engine (Training/Inference)	Framework for Models Management (datasets, trainings, inference)	8/3

In addition to the components of the Software stack, the eFlows4HPC project also plans to provide a set of HPC, HPDA and ML frameworks integrated in the eFlows4HPC software stack through interfaces defined in the project. The following tables provide the current TRL for the mentioned frameworks. In this case **the project aims to elevate to TRL 7** (demonstrated in operational environment) those libraries which are in lower levels.

Data Analytics/Machine Learning/AI

Technology	Short description	Current TRL
ParSoDA	Parallel Social Data Analytic (ParSoDA) is a Java library for parallel data analysis applications	7
DMCF	The Data Mining Cloud Framework (DMCF) is a Cloud system for designing and executing complex data analysis workflows on Cloud platforms	7
dislib	dislib is a distributed machine learning library built on top of PyCOMPSs that provides high programmability and automatic computation distribution in HPC platforms	4
EDDL/ECV	European Distributed Deep Learning library and European Computer Vision library. Both libraries enable using heterogeneous components in the ML process	3
HeAT	Helmholtz Analytics Toolkit (HeAT) is a flexible open-source software for HPDA and machine learning. It provides highly optimized algorithms and data structures for tensor computations using CPUs, GPUs and distributed cluster systems on top of MPI	5
Ophidia	Scientific data analytics framework running over HPC infrastructures, providing parallel, in-memory, server-side data analysis, an internal storage model and a hierarchical data organization to manage large amount of multi-dimensional scientific data	8

HPC simulators/modelling

Technology	Short description	Current TRL
OpenIFS-FESOM2	State-of-the-art coupled climate model used for different climate simulations and applications	8
CMCC-CM2	State-of-the-art coupled climate model used for different climate simulations and applications	8
KRATOS	General Purpose, Parallel Finite Element solver. Contains basic tools for Full Order and Reduced Order models	5
ParMMG	Parallel Mesh Adaptation library	5
Tsunami-HySEA	NLSW Tsunami model with hydrostatic and dispersive versions. Based on a high-order Finite Volume (FV) discretization (hydrostatic) with Finite Difference (FD) for the dispersive component on two-way structured nested meshes in spherical coordinates. Multi-GPU enabled.	7
Salvus	Higher-order continuous Galerkin spectral element modelling of acoustic, elastic and coupled domains in 2D and 3D on CPU/GPU HPC architectures. Developed and maintained by Mondaic	8
ITHACA	ROM techniques built upon a finite volume discretization (implemented in OpenFoam)	5

Finally, another outcome of the project are the workflows for the main target sectors developed in the different Pillars, in this case the project aims to provide these **workflows with at least TRL 5**. As summary, the following table summarizes the TRL for the different **project outcomes**:

Outcome	Description	Target TRL
1	Integrated eFlows4HPC Software Stack	≥ 7
2	HPC Workflow as a Service	≥ 6
3	Integrated HPC, HPDA, ML libraries	≥ 7
4	Pillar's Workflows	≥ 5

1.3.4 National / International Research and Innovation activities linked with the project

The table below reviews some ongoing projects with activities that relate to the project and whose results could be leveraged in the project. Some of the projects listed below include some common components with eFlows4HPC, but none of them has as main objective the delivery of a European workflows platform to enable the integration of HPC, HPDA and AI, that enriches the European software ecosystem and that enables to leverage the performance available in exascale machines. Even more importantly, the proposed HPCWaaS methodology cannot nowadays be found in current projects. In this sense, **the eFlows4HPC proposal is unique in its nature and has the potential to become strategic for the EuroHPC purposes**.

ChEESE https://cheese-coe.eu/ Centre of Excellence that focuses on preparing flagship HPC applications in Earth sciences and natural hazards for the exascale era. It has a strong commitment to building service candidates for the industry and public governance bodies (e.g. civil protection). BSC coordinates ChEESE and INGV, ETH, NGI, and UMA participate in it. **Relevance to eFlows4HPC:** The service prototyping is based upon both existing and novel workflows attached to HPC applications. In particular urgent computing workflows for earthquakes and tsunamis enable a novel usage of HPC resources for very rapid assessment of the impact of natural disasters.
ESiWACE2 https://www.esiwace.eu/ This CoE will enable leading European weather and climate models to leverage the available performance of pre-exascale systems with regard to both compute and data capacity in 2021 and prepare the weather and climate community to be able to make use of exascale systems when they become available. BSC and CMCC participate in ESiWACE2. **Relevance to eFlows4HPC:** Workflows developed in the climate pillar will use models and HPDA tools included in this CoE. The outcome of eFlows4sHPC will be ready to be tested on pre-exascale systems and operational environments.
EXCELLERAT https://www.excellerat.eu/ is a Centre of Excellence that offers cross-cutting support for various engineering sectors, like manufacturing, automotive, energy, aerospace, chemistry, biology and climate, enabling generic application support is a non-trivial task. Within the engineering sector, it provides specialisation on expertise, applications development, targeted training and offered hardware. BSC participates in EXCELLERAT. **Relevance to eFlows4HPC:** The focus of this CoE is related to Pillar I objectives. A win-win situation can be obtained by communicating the pillar results to the EXCELLERAT target groups.
PRACE http://www.prace-ri.eu The mission of PRACE (Partnership for Advanced Computing in Europe) is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realise this mission by offering world class computing and data management resources and services through a peer review process. PRACE is established as an international not-for-profit association (AISBL) with its seat in Brussels. BSC, FZJ and PSNC are members of PRACE. **Relevance to eFlows4HPC:** It has 26 member countries whose representative organisations create the pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level. eFlows4HPC aim to deploy its solution in several PRACE infrastructures, i.e., in MareNostrum (4/5), JUWELS and PSNC resources.
EPI https://www.european-processor-initiative.eu The European Processor Initiative (EPI) is a project whose aim is to design and implement a roadmap for a new family of low-power European processors for extreme scale computing, high-performance Big Data and a range of emerging applications. The goal is to design a processor with vector instructions and specific accelerators with high bandwidth memory access. The EPI processor will also meet high security and safety requirements. Atos, BSC, JÜLICH and ETH are partners of EPI. **Relevance to eFlows4HPC:** Being the target to build the future European exascale infrastructures based on the EPI, it is important to make the codes ready for this processor. eFlows4HPC partners will be able to have early access to the EPI emulator platform with the goal of adapting and optimizing key kernels to this architecture.
EXDCI-2 https://exdci.eu is a joint action of PRACE and ETP4HPC to mobilise the European HPC stakeholders. It aims at the development and advocacy of a competitive European HPC Exascale Strategy by supporting the implementation of a common European HPC strategy, open to synergistic areas including HPDA and AI. Also, it aims at the Coordination of the stakeholder community for European HPC at the Exascale through joint community structuring and synchronisation. PRACE is a partner at EXDCI, and through it BSC, FZJ and PSNC, and CMCC are third parties. **Relevance to eFlows4HPC:** The goal of supporting the convergence of HPC, HPDA and AI which is strategic to EXDCI it is also critical to the eFlows4HPC vision. eFlows4HPC has the potential of contributing significantly to events organized by EXDCI, like the BDEC events.
POP https://pop-coe.eu PoP CoE provides performance optimisation and productivity services for academic and industrial codes in all domains. The offered services are performance assessment and development of proof of concept solutions to improve performance issues and training. BSC is the coordinator of POP, and FZJ is one of

the partners. **Relevance to eFlows4HPC**: The project can benefit from the services provided by POP in its HPC simulation and modelling codes.
ExaQUte http://exaqute.eu aims at enabling Uncertainty Quantification and Optimization Under Uncertainties in complex engineering applications by using complex multilevel Monte Carlo simulations at the Exascale level. With this goal, the project is developing a workflow able to extract parallelism in the MLMC algorithm across samples and levels. One of the workflow frameworks used in ExaQUte is PyCOMPSs. CIMNE is the coordinator of ExaQUte, and BSC and INRIA are partners. **Relevance to eFlows4HPC**: The type of workflow developed at ExaQUte include some of the characteristics sought in eFlows4HPC, such as dynamicity and combination of HPC and data analysis, and its results can be used as input for eFlows4HPC.
AI4EU https://www.ai4eu.eu aims to develop a European AI platform and ecosystem advancing algorithms, tools and resources with the aim to develop AI on demand for Europe and to unify European AI community. Atos and BSC are partners at AI4EU. **Relevance to eFlows4HPC**: The platform proposed by AI4EU is going to serve as a marketplace for AI tools in Europe. eFlows4HPC tools relevant to AI will be registered on the AI4EU platform to leverage the AI4EU ecosystem. Similarly, eFlows4HPC will consider existing tools on the AI4EU platform as components of its platform. Finally, BSC also participates on AI4EU.
Lexis https://lexis-project.eu/ aims to build an advanced, geographically-distributed, HPC infrastructure for Big Data analytics within three targeted pilot test-beds. By proposing innovative technologies and exploiting data available from test-bed partners, LEXIS aims to generate valuable outcomes and improve the efficiency and quality of services provided to different stakeholders involved in the test-beds. Atos and AWI are partners at Lexis. **Relevance to eFlows4HPC**: Lexis is deploying a two-level architecture with orchestration and workflow management. Their experiences on such architecture can be very relevant to the project.
DeepHealth https://deephealth-project.eu. The goal of the H2020 DeepHealth project is to develop two open-source European libraries for ML and computer vision, with the aim to provide an independent European solution to distributed training of health-related applications and to use heterogeneous HPC systems. UPV and BSC are partners. **Relevance to eFlows4HPC**: UPV and BSC are part of the DeepHealth project with the roles of the technical lead of the EDDL development on heterogeneous components and the runtime infrastructure. The EDDL and ECV libraries will be used and adapted to the new use cases of the project. All the knowledge of both libraries will be extended to the project.
HBP https://www.humanbrainproject.eu/ The Human Brain Project (HBP) is building a research infrastructure to help advance neuroscience, medicine and computing. It is one of four FET (Future and Emerging Technology) Flagships, the largest scientific projects ever funded by the European Union. Within HBP, The High Performance Analytics and Computing (HPAC) Platform develops and provides supercomputing, storage, visualisation and simulation technology that can run on supercomputers. FENIX is the infrastructure currently built up by the ICEI project. ICEI is a specific grant agreement (SGA) under the framework partnership agreement (FPA) of the HBP. FZJ, BSC and INRIA are partners at HBP. **Relevance to eFlows4HPC:** HPAC aims at managing complex workflows comprising concurrent simulation, data analysis and visualisation workloads. The results obtained by the HBP project in this activity can be very relevant to eFlows4HPC.
ASPIDE https://www.aspide-project.eu/. Funded under the H2020 call FETHPC-2016-2017. This project provides the design of a new programming paradigms APIs, runtime tools and methodologies for expressing data-intensive tasks on Exascale systems, which can pave the way for the exploitation of massive parallelism over a simplified model of the system architecture, promoting high performance and efficiency. PSNC and Dtok (as University of Calabria) are partners at ASPIDE. **Relevance to eFlows4HPC:** Activities focusing on the implementation of data-intensive applications running on high-performance computing systems will provide software solutions that can be exploited in the workflow-based systems of the project.
MANGO http://mango-project.eu/ and **RECIPE** http://www.recipe-project.eu**.** These two H2020 projects focus on the adoption of heterogeneous HPC architectures and its implications on energy. A large-scale prototype was built in MANGO (coordinated by UPV) made of 192 FPGAs. The RECIPE project (BSC and UPV are partners) works on the provisioning of a hardware abstraction layer for a heterogeneous system. **Relevance to eFlows4HPC:** Both projects will reinforce and guarantee proper adoption of new heterogeneous architectures already worked on FPGAs. The MANGO prototype will be used in the project.
PL-EESM The Pilot Lab Exascale Earth System Modelling (**PL-EESM**) explores specific concepts to enable exascale readiness of Earth System models and associated workflows in Earth System science. PL-EESM provides a new platform for scientists of the Helmholtz Association to develop scientific and technological concepts for future generation Earth System models and data analysis systems. AWI coordinates PL-EESAm and FZJ is a

partner. **Relevance to eFlows4HPC**: Both projects address challenges in the current climate modelling related workflows, in particular data analysis.
PRIMAVERA https://www.primavera-h2020.eu/ PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment (2015-2019). The goal of PRIMAVERA is to deliver novel, advanced and well-evaluated high-resolution global climate models (GCMs), capable of simulating and projecting regional climate with unprecedented fidelity, out to 2050. CMCC and AWI participates to the project. **Relevance to eFlows4HPC**: Some use cases proposed in Pillar II, will leverage models (e.g. CMCC-CM2-HR and -VHR) and end-to-end production-level climate modelling workflows implemented on HPC systems in the PRIMAVERA project.
FocusCoE https://www.focus-coe.eu contributes to the success of the EU HPC Ecosystem and the EuroHPC Initiative by supporting the EU HPC CoEs to more effectively fulfil their role within the ecosystem and initiative: ensuring that extreme scale applications result in tangible benefits for addressing scientific, industrial or societal challenges. **Relevance to eFlows4HPC**: FocusCoE has specific tasks devoted to support the communication between projects/institutions to the CoEs, and its coordinator has agreed to support eFlows4HPC, if funded, on transfer activities to the communities.
ARISTOTLE http://pilot.aristotle.ingv.it and **ARISTOTLE-ENHSP** http://aristotle.ingv.it/ deliver world leading multi-hazard capability to the Emergency Response Coordination Centre (ERCC) of the European Commission, offering a flexible and scalable system that provides multi-hazard services to the ERCC and potentially creates pools of experts in the field of Meteorology and Geophysics that can support the ERCC with regard to situation assessments in crisis situations. INGV coordinated ARISTOTLE and is group leader in ARISTOTLE-ENHSP. **Relevance to eFlows4HPC**: ARISTOTLE provides rapid consultancy to ERCC based on rapid assessments of the potential impact of meteorological or geophysical events (including earthquakes and tsunamis). This service may take advantage in the future of the services developed by eFlows4HPC, as for urgent computing for natural hazards.

1.3.5 Pillar descriptions

1.3.5.1 Pillar I: ROM: Digital twin in Manufacturing

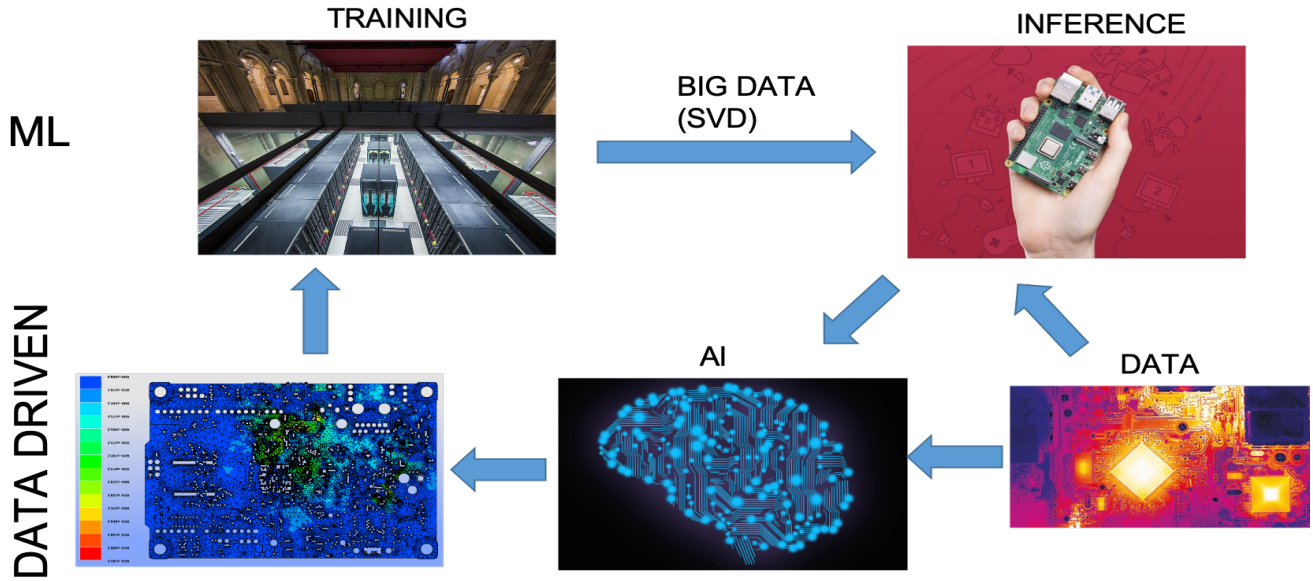


Figure 5: Diagram of Pillar I use cases

The "**Manufacturing Pillar**" will focus on the integration of state-of-the-art adaptive solvers with the eFlows4HPC software stack to allow the digital prototyping of complex manufactured objects. The project will thus aim at the construction of Digital Twins to be used both in the design and in the lifelong deployment of manufactured objects, as envisaged in the framework of the Industry 4.0 vision.
This goal will be approached by combining adaptive solution techniques with Machine Learning (ML) and data mining techniques with the aim of constructing fast Reduced-Order Models (ROM) to be later used at a much-reduced computational expense and eventually deployed on edge-type hardware. Such a goal will be achieved by

advancing both the scope of the simulation technology (to enhance its reach in terms of simulation capabilities) and in the model order reduction technology needed to encapsulate Full Order Models (FOM) into more limited computational environments.
The pillar challenge will be to define, by combining the simulation solvers, ML and other components of the eFlows4HPC software stack, a simulation framework to be demonstrated in relevant scenarios as defined by the industrial partner.
A specific objective of the team will also be to provide first class integration with the Artificial Intelligence (AI) modules available in the project. This will allow two-way integration between the simulation and the AI modules, and will enable on one side, the integration of the simulation tools as a source of training data, and on the other side, the use of AI models in steering computational results.
The participation of Siemens as a project partner will be key to guarantee a real-life impact on selected use cases. As an example, we devise the application of the proposed framework to the construction of a Digital Twin for high power transformers. This will enable operational control of such items (which would be of high industrial interest), will help toward their design optimization, and will enable the construction of deterministic models to relate sensor data (only available at certain spots where the sensors can be eventually built into the machine) to the behaviour in critical areas for which direct observation is currently impossible.
At the same time the integration of such capabilities in the proposed HPC Workflows as a Service framework will open new possibilities of outreach and will eventually enable to achieve a project sustainability well beyond the time frame of the project.

Data assets description
The basis of the "Manufacturing pillar" is the solution of a large number of training cases, which need to be mined to extract reduced order bases. The practical implication is that the bulk of the data (intended as simulation results such as displacements, temperatures, stresses or finite element data as required for hyper-reduction) will be produced and consumed within the same workflow implying that a large percentage of the required data allocation will be short lived. As an estimate, a typical training scenario for a model containing one million elements will require the storage of about 30 million double-precision (DP) numbers. Considering storing 1000 scenarios this implies a total allocation of about 30.000 million DP numbers, which correspond approximately to 240GB. The long-term allocation, required to store the result of the model reduction will be typically much reduced, of the order of 20 million DP numbers or around 160MB (per model).
Considering that larger models and multiple simulations can be targeted, we estimate a total required allocation of the order of 5TB, even though a larger scratch space allocation should be accounted in order to provide a buffer for the workflow space requirements.
The level of openness of the data generated during the project can vary depending on the specific model considered. Training data for benchmark scenarios will be considered for publication if deemed relevant. Industrial models will be kept as confidential according to the industrial partner's requirements. **The data sets used in this pillar do not include personal data.**

1.3.5.2. Pillar II: Dynamic and adaptive workflows for climate modelling

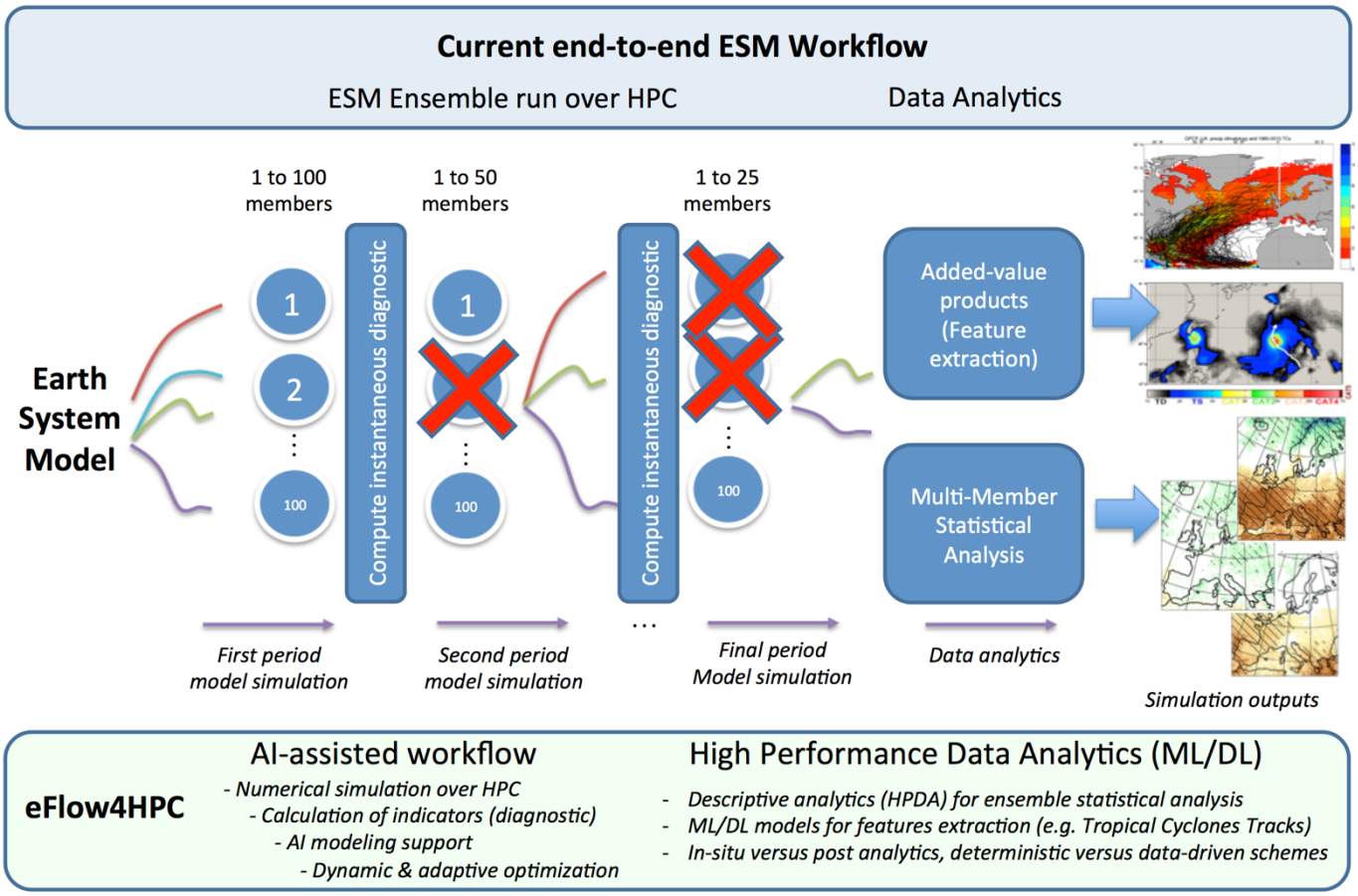


Figure 6: Diagram of Pillar II use cases

Pillar II will develop innovative adaptive workflows for **climate modelling** making an efficient use of the computing resources by performing a dynamic pruning of the members in climate ensemble simulations. We propose a novel approach where the number of members starting the simulation is higher (hundreds of them), and the selection is done during the simulation in regular steps. In a determined step, without stopping the simulation, an index (diagnostic) is computed and the members are classified following this index. Those members that are under a threshold are discarded and stopped, while those that are still "valid" can continue until the next step. This approach is useful for minimizing present-day bias and obtains more realistic ensemble mean and control the ensemble spread. It can be also used for parameter optimization and for attempts to closely reproduce some specific past climate conditions or events. Resources freed by the discarded members can be freed, and therefore, the simulation has smaller energy consumption. Pillar II will also focus on feature extraction to infer (besides canonical climate variables) also added-value products. In this respect, the Tropical Cyclone (TC) track will represent a notable case. Such data analytics task will be (i) initially tested in a multi-model scenario with data from the Coupled Model Intercomparison Project Phase 6 (CMIP6[vi]), focusing on very high-resolution models' output (i.e. HigResMIP) and then (ii) integrated into ensemble ESM operational workflows. TC tracking analysis can be very challenging due to the large amount of data involved, heterogeneity of data and complexity of the analysis/processing required. Different tracking methods are available in the literature and new emerging approaches are investigating the use of ML/DL techniques. eFlows4HPC will orchestrate into the same end-to-end scientific workflow, in a seamless, intelligent and transparent way, climate models, ML/DL and HPDA tools. It will also integrate and compare data-intensive versus data-driven approaches taking into account efficiency dimensions (e.g. time-to-solution, energy consumption). In terms of outcomes, while the technical solution will demonstrate the feasibility of the proposed approach (innovation) on large and complex production-level ESM

[vi] https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6

workflows, the scientific one will regard added-value datasets for a wide spectrum of sectors directly connected with climate change (societal impact).
Data assets description Pillar II will involve climate simulation datasets from one of the largest community experiments in the climate domain: the CMIP6 and in particular it will consider high-resolution datasets from the HighResMIP[3] for the CMIP6 experiment. In general, CMIP6 datasets consist of NetCDF[vii] (files with a single output variable along with coordinate/grid variables, attributes and other metadata) from a single model and a single simulation (i.e., from a single ensemble member of a single climate experiment). There is flexibility in specifying how many time slices (samples) are stored in a single file. A single file can contain all the time-samples for a given variable and climate experiment, or the samples can be distributed in a sequence of files. The metadata is constrained by the CF convention (NetCDF Climate and Forecast (CF)[viii] Metadata Convention) and as specified in the CMIP6 tables. The output files are written through the NetCDF API[ix]. CMIP6 data publication is still on-going and it is expected to deliver to the community about 20PBs of (open) data. **The data sets used in this pillar does not include sensitive data.** Data generated by coupled OpenIFS-FESOM2 and CMCC-CM2 model simulations will be used for evaluation of dynamic pruning of the members in climate ensemble simulations and for Tropical Cyclone (TC) track analysis. The amount of data depends on the selected horizontal resolution of atmosphere and ocean components as well as on the data storage frequency. The higher the resolution, the more often, usually, one has to store the data to investigate fast events reproduced by high resolution model. We will begin with ensemble experiments with modest resolutions in atmosphere and ocean and produce data of order of several Terabytes per simulation. Later, for tasks like TC track analysis we plan to target very high resolution in both atmosphere and ocean components, that can increase the output dataset size to order of several hundreds of TBs. In terms of data model related to the simulation output, it is important to consider that in OpenIFS-FESOM2, ocean model uses unstructured grid, while grid of the atmospheric component is structured, whereas in CMCC-CM2 both ocean model and atmospheric components use structured grids. The output files are written through the NetCDF API.

1.3.5.2. Pillar III: Urgent computing for natural hazards

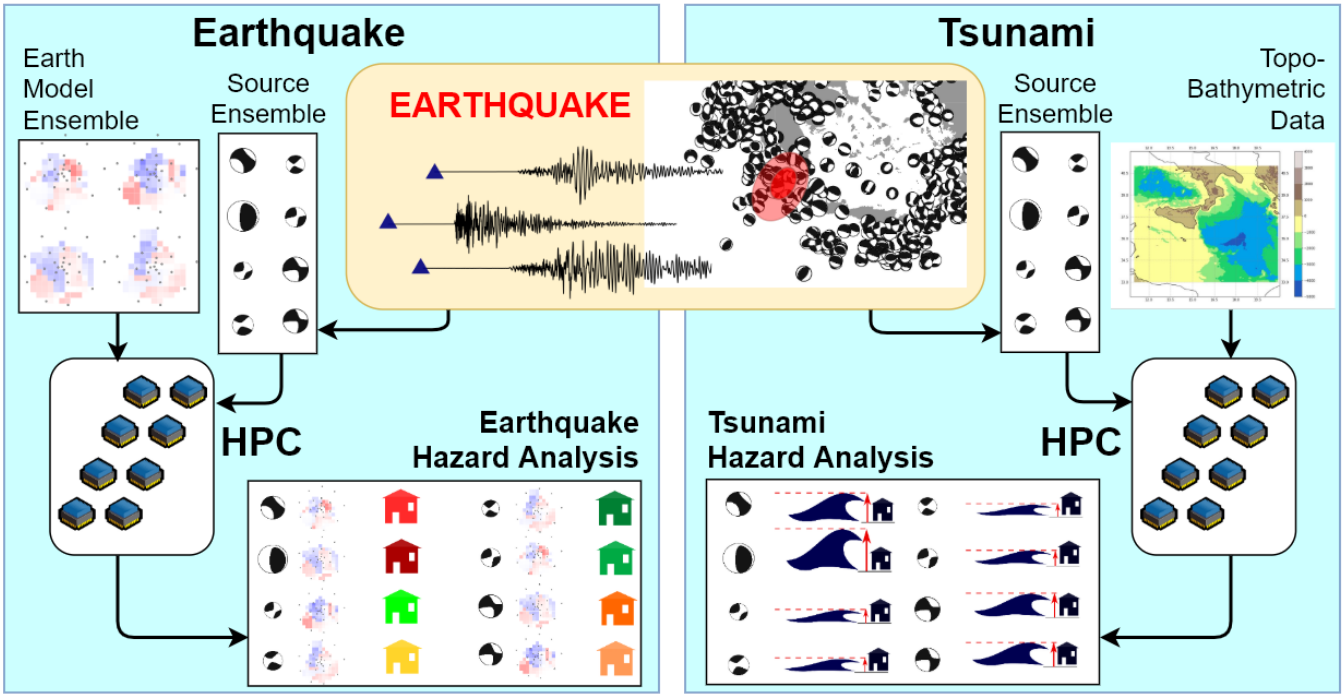


Figure 7: Diagram of Pillar III use cases

[vii] https://www.unidata.ucar.edu/software/netcdf/

[viii] http://cfconventions.org/

[ix] https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_apis.html

This pillar is related to the modelling of **natural hazards,** in particular the occurrence of earthquakes and their associated tsunamis, urgently after such an event is detected. The main idea is to exploit the high resolution attained by numerical simulations to rapidly provide a deeper insight into the potential consequences of the event. Key to the workflow's success is improving our capacity to use significant HPC resources with little to no queuing time, as well as the ability to incorporate existing data seamlessly into the process. Our capacity to correctly quantify uncertainties is critical, as is the capability to deliver timely and relevant results to end-users: particularly to key stakeholders in rapid response to the disaster relief efforts. We will add capabilities related to data integration to the existing HPC workflows, improve the workflow robustness, and propose policies for urgent access to supercomputers.

The most important challenges in improving the quality of the procedures are: 1) obtaining high-resolution Earth models (e.g. maps of the properties of the subsurface, bathymetry); 2) accurately estimating the impact of parameter variations in the outcome of simulations, i.e., the parameter sensitivity; and 3) ensuring fast and reliable results with urgent access to computational resources and smart management of all workflow components. The data assimilation capabilities will allow dynamic evaluation of the numerical simulations against observational constraints. This requires interaction with the workflow's job pool and, potentially, dynamic allocation of computational resources (i.e. resource elasticity). Our workflows will be evaluated against already existing solutions, which employ simpler non-physical assumptions and are currently in service at civil protection organizations.

Data assets description

The pillar will involve multiple sources of data of different nature (all the following data is public and open). First, parametric data for earthquakes (e.g. localization, magnitude, moment tensor), partially available through web services (FDSN event WS[x], INGV ONT[xi], etc.). The uncertainty in these parameters will be also based on other non real-time datasets, such as global and regional Centroid-Moment-Tensor (CMT[xii,xiii]) or fault databases (Database of Individual Seismogenic Sources (DISS)[xiv]). Waveform data for earthquakes to extract Peak Ground Motions (PGM) from broadband and accelerometric networks, such as EIDA[xv], IRIS[xvi], ESM (Engineering strong motion[xvii]), or local networks (e.g., for Italy, Rete Sismica Nazionale - RSN[xviii] and Rete Accelerometrica Nazionale[xix]). Waveform data for tsunamis to exact tsunami intensities, like tide-gauge data from the IOC Sea Level Station Monitoring Facility, accessible through web services[xx] or accessible through GTS (WMO), or provided directly from INGV converted into miniseed[xxi] (also obspy) format, including also data from other sources (e.g. local networks), as well as other data like DART Buoys data[xxii]. Open source topo-bathymetric data will be used as input for setting up the tsunami simulations. To this end, data sources such as GEBCO[xxiii] or SRTM[xxiv] will be employed and used to grid the necessary input files for the tsunami simulations.

The databases themselves can hold datasets several hundreds of TB in size, albeit typical queries should require very small subsets, in the order of tens of MBs. In the construction of 3D Earth models, thousands of such queries might be required. The models themselves can be several GBs in size, depending on the required grid resolution. The databases are typically open although the output of our simulations, for real earthquakes and tsunamis, is potentially sensitive for public opinion and hence requires careful handling.

[x] https://www.fdsn.org/webservices/

[xi] http://terremoti.ingv.it/en/webservices_and_software

[xii] https://www.globalcmt.org/

[xiii] http://rcmt2.bo.ingv.it/

[xiv] http://diss.rm.ingv.it/diss/

[xv] https://www.orfeus-eu.org/data/eida/networks/

[xvi] http://ds.iris.edu/ds/nodes/dmc/data/

[xvii] https://esm.mi.ingv.it/

[xviii] http://terremoti.ingv.it/instruments/network/IV

[xix] http://ran.protezionecivile.it/IT/index.php

[xx] http://www.ioc-sealevelmonitoring.org/

[xxi] https://ds.iris.edu/ds/nodes/dmc/data/formats/miniseed/

[xxii] https://www.ndbc.noaa.gov/dart.shtml

[xxiii] https://www.gebco.net/data_and_products/gridded_bathymetry_data/

[xxiv] https://www2.jpl.nasa.gov/srtm/

The following table summarizes the preliminary requirements from the Pillars with regard to the use of HPC, ML and data analytics and to workflows' characteristics.

Pillar	Use of HPC	Use of ML and data analytics	Requirements on workflows
Pillar I	MPI/OpenMP for training. MPI/OpenMP/Accelerators for basis extraction. OpenMP/Accelerators for inference	Use of distributed ML algorithms for the extraction of data using SVD-like approaches and hyper-reduction	Dynamicity, adaptivity, resiliency, openness of data/workflows, ease-of-use
Pillar II	Climate models, post-processing tools and data analytics applications (MPI/OpenMP)	Feature extraction (like TC track) and ensemble analysis. Used in data-intensive and data-driven scenarios, with respect to both post- and in-situ analytics use cases. Data/products will be stored in NetCDF format and the CF convention will be adopted to promote the processing, sharing and reuse of NetCDF files	Dynamicity, adaptivity, resiliency, usability, transparency, openness of data/workflows for further re-use
Pillar III	Heavy use of CPU and GPU models	Training a neural network to compute fast shakemaps out of earthquake parameters. Useful in sensitivity studies	Redundancy, elasticity, resiliency

1.3.6 Gender Analysis

In accordance to the EU policy the project consortium is committed to foster the principles of gender mainstreaming and every effort to ensure that the project and related activities contribute to the promotion of gender equality, as well as that none of the activities within the project contribute to gender inequality.
eFlows4HPC consortium is aware of the importance of gender balance, and will promote this aspect during the whole project lifetime, with the following objectives that underspin the gender action plan:

a) to ensure that women and men have equal opportunities to participate in the various parts of the project;
b) to contribute to reduce inequalities between women and men, having both women and men in the different management roles, which is already the case with a female project coordinator;
c) to take active measures to ensure that women scientists are well represented in the consortium, both on the ICT side and on the Pillars' implementations, and that women scientists lead elements of the project, which is already the case in the team designated by the different partners to do the work, as presented in section 4.1; and
d) to offer opportunities for mobility within the project that take into account the different needs of women and men in order to enhance participation by women scientists.

1.4 Ambition

1.4.1 Workflow technologies for the convergence of HPC, HPDA and AI

The project aims at defining methodologies for the development of workflows, where different aspects are integrated: data management and data analytics, high-performance computing and machine learning processes. However, in current best practices we find all these elements in separate components and environments.
Workflows provide a high-level paradigm for specifying the logic of an application, hiding the low-level details that are not fundamental for application design. They are also able to integrate existing software routines, data sets, and services in complex compositions that implement scientific discovery processes. A main issue in scientific workflow systems is the programming structures they provide to the developer who needs to implement a scientific application[4].

Various proposals have been formulated for the development of scientific workflows. Some are based on graphical interfaces[5,6,7], on textual interfaces[8,9,10,11], or on programmatic interfaces[12,13]. An interesting aspect of the scientific workflows' environments is that each scientific community seems to stick to one or another solution. For example, Galaxy[28] has been adopted by the ELIXIR[14] life science research infrastructure as its main workflow environment, while Cylc[86] was selected among others by the Earth Science community. An important component of workflow environments is their runtime or engine, which is responsible for coordinating the execution of all workflow tasks, scheduling them in the available computing resources, transferring the data between distributed storage systems, monitoring the execution of the tasks, etc. As with the interface, the runtime can be different from one environment to another, from very simple to more sophisticated ones, sometimes implementing different techniques to optimize the execution of the workflows and reduce the amount of required data transfers, for example. The support for the management of metadata is not common, but can be found in Galaxy.

Traditionally, workflow systems did not entail the possibility of supporting parallel or HPC tasks (tasks that run in parallel in large HPC infrastructures). Typically, HPC applications are developed using the Message Passing Interface (MPI[15]) programming model, which is the de-facto standard for this type of applications. It is based on the idea of having a large number of concurrent processes that exchange messages with the objective of solving a large problem in a cooperative way. Currently, MPI is combined with other approaches to exploit the concurrency inside the large and fat HPC nodes. The most popular approach for this is OpenMP[16]. An additional complexity for the application developers of is the appearance of accelerators such as the graphical processing units (GPUs) which require specific programming environments, like CUDA[17]. HPC programming models tend to be quite low level and require a lot of effort from the application developer.

We find differences with respect to the programming models for big data applications with regard the way workflows and HPC applications are programmed. The most common approach for big data applications is Spark[18], an extension of MapReduce[19]. MapReduce is based on two operators, map and reduce. Developed code need to use both appropriately. Indeed, Spark has generalized the concept and offers up to 80 operators to assist code development. While both approaches can run in distributed computing platforms, the differences with traditional HPC codes written in MPI are significant and both communities follow different approaches.

The way Machine Learning applications are programmed is closer to big data ones, i.e., in terms of the programming languages that are used and environments. There are multiple libraries and environments available, most of them offering a Python interface[20,21,22,23]. In some cases, they are built on top of big data environments, like the MLlib library[24] that runs over Spark. A remarkable and emerging case is the European Distributed Deep Learning (EDDL) library, which is being developed in the H2020 DeepHealth project[25]. EDDL uses C/C++ interfaces as well as Python wrappers.

To address the main challenges in science and engineering applications, workflow programming frameworks must include adequate abstractions for data representation and concurrent processing orchestration. Besides the amount of data accessed and analysed by the workflow tasks, the output data produced by each workflow node or stage needs to be stored and annotated. This enables its reuse in later further executions or new workflows. A crucial, yet often neglected aspect of modern data-driven science is the availability of the relevant data on the processing site. For that, data residing in different sources must be brought together, integrated, cleaned, and transformed. Given the far-reaching distribution of such resources and the fact that data is often updated, the process of making data available for computation becomes a repeated and tedious task. For the sake of reproducibility, the task of data preprocessing must not be conducted manually but rather formalized in terms of a processing script or workflow[26]. The availability of huge amounts of data, or Big Data, originated a data-centric science based on the intelligent analysis of large data repositories to extract the rules hidden in raw data coming for laboratory experiments of from physical phenomena. Workflows offer a programming support for researchers in implementing discovery tasks by coding in a computational graph an entire scientific practice or methodology that is too complex or impossible to implement in a laboratory. Moreover, many scientific workflows have to operate over heterogeneous infrastructures with the added possibility of failures. Therefore, dealing with such computing infrastructures in a more coherent way, is another important challenge to be addressed and solved.

Progress beyond the state of the art

The current methodologies available to develop scientific workflows:

a) Do not fulfil the requirements of increasingly complex applications, which require novel methodologies that support in a holistic workflow HPC simulations or modelling, data analytics, and machine learning.

b) Are not able to get the best of the complexity of the underlying infrastructure, which is distributed in nature. The infrastructure we consider is composed of a large number of nodes with new types of processors with multiple cores (including accelerators such as GPUs), new storage devices that have the potential to change the way data is stored by applications, and connection to external instruments, edge devices and cloud storage as sources of data.

c) Do not include techniques for dynamically adapting the execution of workflows on computing platforms.
d) Do not provide a fully integrated approach to the management of both HPC and Big Data analytics requirements for providing data-oriented frameworks easy to be extended with additional functionalities related both to HPC tasks and data analytics techniques.
e) Neglect, or do not address to the full extent, the challenge of making the required data available for processing on time, in expected format, and quality.

eFlows4HPC will do progress beyond the state of the art proposing: ground-breaking solutions for interfaces for the development of scientific workflows that will integrate HPC applications, machine learning and big data; novel intelligent runtimes that will optimize the execution of the workflows, reducing the space exploration and the energy required for its execution; innovative tools for workflow modelling; always leveraging on European solutions. It will also provide a Data Logistic Service to fuel the data-intensive workflows with relevant, up-to-date data in a reproducible and transparent manner and Data Catalogue to list potentially relevant data sources.

1.4.2 HPC Workflows as a Service

One of the main issues that an HPC workflow developer and system administrator has to deal with is the installation and deployment of the workflow dependencies. Due to the widespread number of compilers, library versions, and their incompatibilities, every time a user wants to deploy a new workflow in a supercomputer, she has to check the installed dependencies, and install the missing ones taking into account the libraries and compiler versions in order to detect possible incompatibilities. To mitigate these issues, we can find some tools such as Spack[27] or Easybuild[28] which provide mechanisms to deal with these issues and automate the installation process of new software in HPC environments. However, they still require an expert HPC developer to create the packages or recipes for these tools and validate they are working for each supercomputer.

In Cloud environments, virtualization and container technologies have simplified the portability of complex applications. Hypervisors such as KVM[29] or container engines such as Docker[30] allow running processes in customized environments on top of computing nodes. These environments can be customized as normal computers and can be saved in images, which can be easily copied to other nodes where the same process can be executed with the same environment. The main barriers to apply these technologies in HPC is rooted with the requirement of hypervisors and engines to run in privilege mode (root access) with the security consequences that this implies, and the integration of images with some specific HPC hardware such as fast interconnects drivers. Singularity[31] is a container engine which tries to overcome these issues by not requiring privileged user mode to run the container and allows the direct access to the host drivers to benefit from the special HPC hardware. Apart from this, Cloud Computing also provides service-oriented abstractions called *Everything as a Service* where a set of services is offered depending on the usage model. One of the latest proposed service models is the *Function as a Service (FaaS)*. This service enables users to execute functions in the Cloud in a transparent way with a simple REST API call and without having to deal with the entire deployment, configuration and execution management overhead. The *FaaS* platforms, such as the commercial AWS Lambda[32], Google Cloud Functions[33] or the open source approaches like OpenWhisk[34], are in charge of managing the different function executions, allocating the computing resources when required, deploying the function software, getting the input data and storing the output results.

Progress beyond the state of the art

The eFlows4HPC project will extend the FaaS model to provide the same simplicity and transparency to the execution of complex workflows in federated HPC infrastructures. Therefore, HPC users will be able to execute a workflow by just performing a simple REST API call. The eFlows4HPC software stack will manage all the complexity of deploying, configuring and executing the workflow in the HPC infrastructure. To achieve this goal, the project will leverage the TOSCA Cloud automation and container technologies. In addition, these technologies will allow setting up registries of workflows, i.e., catalogue of re-usable and ready to deploy/execute workflows.

1.4.3 Model repositories

Data-intensive science produces large amounts of data. This data is in turn analysed to produce new scientific insights. This can be done manually (e.g. visualizations or statistical evaluations) or, alternatively, machine learning (ML) approaches can be used to identify patterns and trends with various algorithms. The increasing volumes of data caused increasing popularity of the later methods. ML uses (large amounts of) training data to produce a mathematical model that can be used to make predictions, classifications, and decisions. The training process can be very laborious due to the size of the data and often requires special hardware (high throughput systems, GPUs, etc.). We can observe constant progress in algorithms used for machine learning (e.g. deep learning), and improvement in the performance of the produced models.

As Howison and Bullard[35] pointed out software citation suffers under similar problems as those known for data citation. The challenge is to apply FAIR principles (findable, accessible, interoperable and reusable) to software and there are some first attempts on achieving that. In this project, however, we want to address the challenge of sharing

not the software but instances of the trained models, i.e., the results of laborious training. There are not many solutions for this. Google owns a patent on "Machine learning model repository"[36] and there are a number of proprietary implementations: Google TFX, Facebook FBLearner and Uber Michelangelo.

Progress beyond the state of the art

The project will progress beyond the state of the art by providing a model repository where trained ML models will be stored with additional metadata describing the parameters used for training and performance metrics of the created model. The repository will also provide a deployment function, allowing reuse of the models, e.g., classification of new data. It should be stressed that, although specialized high-performance hardware is often required to train the model, the subsequent prediction and classification (inference) can be done with low-end resources. The concept will be tested on the use cases provided by the Pillars.

1.4.4 Heterogeneous support for HPC

With the end of Dennard's scaling[37] and the ever-increasing demands for higher performance from conventional HPC numerical codes as well as emerging applications (such as, for instance, ML), in the last years the computer architecture scene has become much more diverse. As a result, we can now find processors that aim to augment the floating-point operations per second and Watt (FLOPS/W) following different paths, for example: 1) exploiting very wide SIMD units, as in the Fujitsu (ARM) A64FX, ranked in the top position of the last Green500[38]; 2) integrating a large count of cores, as in PEZY Computing's PEZY-SC2, ranked #2 in the same list; and 3) using more conventional, yet heterogeneous designs combining general-purpose processors from IBM (Power9) or Intel (Xeon) with NVIDIA's GPUs, present in positions #3-10 of that list. The recent trends in HPC confirm that a hybrid architecture combining CPUs, GPUs, and even specialized accelerators has become the preferred node type for a large range of workloads of interest for HPC and data centres, including ML, Big Data, and scientific simulation.

In addition to well established accelerators like GPUs, reconfigurable hardware devices like FPGAs have also been gaining popularity over the last few years in HPC environments, as obviously confirmed by successful projects such as Catapult[39]. FPGAs are nowadays offered as a component by companies like Alibaba when hiring virtual machines in their cloud systems. Moreover, in an attempt to bound response time to users and minimize power consumption when providing DL-based services, Google has developed a new kind of accelerator, the Tensor Processing Unit (TPU)[40], delivering substantial higher throughput and more favourable performance/watt ratio compared with contemporary CPUs and GPUs.

Progress beyond the state of the art

eFlows4HPC will promote the adoption of heterogeneous architectures based on GPUs, FPGAs, and custom accelerators in European industry, represented by the three Pillars' use cases. Moreover, the application of Deep Learning technologies to accelerate parts of the Pillars' applications will be exercised in the project and represents an innovation achievement. eFlows4HPC will incorporate additional emerging solutions at hardware level. Reconfigurable hardware, such as FPGAs, is getting maturity and is being considered as a new component to be used in HPC and BigData systems. The target applications, by convention, run on full 64-bit floating point precision. The project will explore more energy-efficient implementations by trading off intermediate precision versus performance/energy while maintaining the precision of the final results.

1.4.5 Workflows for ROM in Manufacturing

The use of simulation in the design phase is currently pervasive to the point at which simulation effectively replaces experimentation in many areas. The next frontier[41] is the generation of Digital Twins, which integrate different simulation technologies in order to construct a digital representation of the object of interest[42,43]. Such Digital Twins[44] can be fed with measurement data and used to predict the future behaviour, thus opening the door for improved lifetime management strategies. Think for example of the early prediction of impending failures, or the experimentation of extended usage limits.

State of the art tools for the simulation of complex industrial problems have several issues. Standard FEM solvers may fail in exploring large training spaces, and thus need to be complemented by robust mesh adaptation strategies and novel immersed solution capabilities. Moreover, the solution workflow needs to be reworked so to guarantee full interoperability between the solution strategies and ML frameworks.

On the solver side, the team will tackle the challenge by combining the use of state-of-the-art solution approaches, as for example the "Shifted Boundary Method" [45,46,47] or other in-house alternatives[48] with aggressive mesh adaptation techniques[49], an approach that will enable the fully automatic solution of the problems of interest.

Such detailed simulations, known as "Full Order Models" (FOM), are still too expensive to enable interactive usage; a feature needed in view of deploying digital twins on small form factor devices as the on-board computers of the manufactured products. Achieving the goal thus implies embracing the use of model reduction techniques, which offer the possibility to provide predictive response at a vastly reduced computational cost. From a technical point of

view, Model Order Reduction is one of the most relevant mathematical techniques for the development of applications, such as in data science, HPC, and web computing, which rely on fast or real-time computing.
Reduced-order models (ROMs) consist into simplified mathematical models derived from the full set of parametric partial differential equations (PDE), which govern the physics. See for example [50,51] for applications in the context of CFD using different discretization approaches.
In practice, a reduced-order model workflow is based on two main steps:

- in the first step, generally referred as "**Offline**[52]" (or "**training**" in the AI community), the space of parameters is sampled in a proper way. Highly accurate Full Order simulations for the selected sets of parameter values are run in order to have a comprehensive description of the solution manifold. At the end of this step, Reduced Order Bases are generated starting from the set of full order solutions.
- in the second step, known as "**Online**" (or "**Inference**" in AI), the reduced-order bases obtained at the end of the first step are used for simulation using a new set of parameter values, different from the ones used during the training.

The "**Offline**" phase is computationally expensive. The execution of the training campaign, which may need to be iteratively refined based on the results of the ROM training, needs to be fully automated and thus requires combining state-of-the-art solution technologies with adaptive workflows. Furthermore, the extraction of patterns from the training data constitutes a computationally expensive data mining procedure, generally based on the use of the *Singular Value Decomposition* (SVD). To alleviate the intrinsic computational burden of the SVD, one may resort to distributed *randomized factorizations*[53,54]. Alternative techniques substitute the SVD by the solution of a succession of generalized eigenvalue problems to identify linear vibration modes, enhanced by the so-called *modal derivatives*[55]. The simulation of highly nonlinear problems might also exceed the practical representation capabilities of SVD (or modally enhanced) bases. In such cases, nonlinear dimensionality reduction methods, such as the Isomap algorithm[56], may be more suitable to unravel and extract the dominant patterns. This implies moving to more complex, non-invariant, bases taken as the tangent space to the solution manifold[57,58].
The Model reduction is finally completed by the choice of optimized integration techniques, an approach known as "**hyper-reduction**". This step may be carried out by interpolation of (nodal) nonlinear terms, as in the EIM and DEIM methods[59]; or by efficiently integrating the nonlinear term, using an optimally selected subset of elements as in the (ECSW)[60] method, and in the *Empirical Cubature Method*[61].
The "**Online**" phase on the contrary is computationally inexpensive, thus allowing fast predictions to be performed in computational constrained systems such as industrial controllers, on-board computers (think of Raspberry PIs) or on common smartphones/tablets. This allows true model-based control of operating assets, or even real time decision support, as the operator will be able to pre-calculate different operation scenarios and thus to make better informed decisions. The price paid to achieve such result consists in a slight, controllable, reduction in the accuracy of the results with respect to the ones obtained by FOM.

Progress beyond the state of the art

Overall, the objective of the manufacturing pillar will be to **integrate adaptive solution capabilities and ROM techniques within a single eFlows4HPC workflow**. While a number of efforts exist to accomplish the construction of ROM (particularly in specific fields), we are not aware of general purpose open-source developments specifically targeted at next generation HPC, a gap we aim at closing with the current project. The team will also work on the development of novel model reduction techniques specifically tailored for full order solvers, which consider automatic mesh adaption and generation. Some preliminary works[62,63] have been conducted for the case of linear problems and one of the aims of the present project is to extend these efforts to a general non-linear and more complex scenario.
The pillar vision is completed by the integration of the simulation capabilities within a workflow which includes AI capabilities, which will permit making "intelligent" use of the simulation, for example in steering the manufactured system's response (think of avoiding overheating for a thermally constrained problem), or to eventually predict impending failures.

1.4.6 Workflows for Climate modelling

End-to-end Earth System Modelling (ESM) workflows consist, by nature, of different steps including numerical simulations, data post-processing, as well as data analytics and visualization. Even though they represent different pieces of the same picture, their ***seamless***, ***intelligent*** and ***efficient*** integration into HPC environments still needs to be addressed at different levels to become a reality.
In general, full workflow management systems for climate modelling are not yet well established. In most of the cases, scientists develop their own scripts and command line interfaces to address operational needs and integrate different components of their workflows. Such an approach is clearly difficult to maintain/extend; it poorly addresses re-usability; and it does not introduce any optimization, decoupling or abstraction layer. Moreover, due to its nature

it needs the workflow logic to be included as part of the workflow scripts requiring ICT expertise from the scientists and becoming itself an error-prone development. Such an approach does not represent an adequate solution with the current size of data as well as with the need for more workflow components (HPDA/ML/DL), and it will definitely fail in the near future with very high spatio-temporal resolutions, even in the context of a single climate simulation.
Over the last two decades, from a workflow management perspective, a lot of tools (general-purpose or domain-specific) have been used for research purposes to address workflow needs in distributed environments or at the data centre level. However, none of them have been significantly and pervasively (from a community adoption perspective) adopted into operations with a large consensus (production scenarios at different centres have mainly leveraged site-specific solutions).
As an exception to that, over the last few years the Cylc[64] workflow tool has gained popularity, increasing interest and adoption in the climate community in terms of workflow management support for weather and climate simulations. However, the focus of Cylc is still more on the numerical simulations (rather than the full end-to-end ESM workflow or the integration with HPDA/ML/DL components); it does not offer a data science-oriented programmatic layer; and it only provides a static support in terms of workflow runtime execution without addressing optimizations based on energy consumption or learning from previous workflow executions. The same is true for recent attempts of the ESM community to abstract some parts of the climate modelling workflow using Python based tools (https://www.esm-tools.net/)
The integration with data analytics aspects as well as the multifaceted nature of ESM workflows in the climate domain has also raised the need to support multiple levels of workflow management. As an example, two-level workflow approaches (i.e. Ophidia/Kepler[65 66], Ophidia/COMPSs[67]) have demonstrated the ability to properly address both the horizontal extension/distribution of a workflow and the vertical execution of a single step (equivalent to a micro-workflow) on a specific component (i.e. data analytics). Such requirement becomes a relevant challenge for end-to-end ESM workflow scenarios, where HPDA, ML/DL components would be also involved.

Progress beyond the state of the art

The current methodologies available to develop scientific workflows:

a) Do not seamlessly, transparently and efficiently integrate climate simulations, HPDA and ML/DL components into the same end-to-end ESM workflow.
b) Are not able to be adaptive and dynamic and mostly perform a brute force approach, thus preventing to save computation/energy costs as well as to reduce the total execution time.
c) Do not take advantage of or properly manage the complexity and heterogeneity of the underlying infrastructure, thus negatively impacting on efficiency.
d) Do not sufficiently address FAIR[68] principles and open science.

The project will do progress beyond the state of the art proposing: a more holistic approach with respect to end-to-end climate workflows; intelligent approaches to optimize workflow execution and reduce computational cost; transparent access to the underlying infrastructure and open repositories to address reuse of workflow templates, but also a more reproducible science. For climate modelling this will improve the confidence of scientific results by minimising unnecessary computations and reallocation of resources to more relevant tasks. Inclusion of HPDA and ML/DL into the climate modelling workflow will allow performing exploratory data analysis on very high resolution/very large ensemble model data. Inclusion of climate modelling workflows in HPCWaaS will democratize access to climate model simulations and data analysis for scientists that have no previous experience with HPC systems.

1.4.7 Workflows for Urgent computing for natural hazards

Early decisions in rapid response to earthquakes have to be based upon interpretations of the best, yet often limited, data available at a given time following the earthquake occurrence to estimate the severity of shaking and, potentially, of the subsequent tsunami. A combination of data analysis and simplified models based upon early records produces shake maps which can be employed to assess losses (e.g. in the insurance industry) and direct relief measures (e.g. civil protection). While numerical simulations can model the physical phenomena with great accuracy, computed outcomes are very sensitive to uncertainties in the source characteristics (earthquake type, size, structure, location), which may result in large variability of the impact figures. Such input data sensitivity, possibly enhanced by limited data availability, and the inherent computational cost of high-fidelity geophysical simulations have rendered numerical simulation as an analysis tool to study both past and hypothetical future earthquakes and their associated tsunamis.
In practice, the increase in computer power makes it far less difficult to run accurate simulations of these natural disasters in a useful time frame. Tsunami simulations can be efficiently run in a single GPU faster than real-time[69] and 3D earthquake simulations, based upon explicit time integrators, are among the most scalable applications in HPC. In addition, the possibility of addressing uncertainty by means of exhaustive exploration of the parameter space

near a guess of initial conditions[70] (i.e. early source estimates) results in a problem with computational demands that can be handled in current-generation large clusters. One of the primary limitations to the application of numerical simulations of earthquakes and tsunamis for disaster resilience is thus the means to access large HPC resources at very short notice: i.e. *urgent supercomputing*.

Urgent supercomputing for the aforementioned natural hazards involves complexities at many levels. From a **technical** perspective, on the one hand, the frameworks required to manage the data intake, analysis, simulation pre-processing and post-processing are very complex. Furthermore, given the critical value of the results, we need to ensure a certain quality of service and resilience to system failure, so that results can be available at key decision centres. Last but not least, there must exist access policies to HPC clusters that reflect on the urgent nature of such simulations, which differ radically from current resource access policies at public supercomputers within PRACE (or EuroHPC). From an **integration** perspective, the workflows developed must connect seamlessly with the needs of the stakeholders in disaster management, who already have tight communication and decision protocols. From a **scientific** perspective, the results should provide the state-of-the-art in terms of fidelity to the recorded data and compatibility with physical constraints.

Progress beyond the state of the art

The current methodologies for early warning or fast impact assessment:

a) Do not look beyond principal parameters (hypocentre and magnitude).
b) Can not employ full physical simulations, rather simplifications or a discrete set of precomputed scenarios.
c) Ignore uncertainties specific to the early information sources used.

Current practice in the context of **tsunami** early warning and/or rapid post-event assessment is either to resort to very conservative (pessimistic) tsunami impact estimates using relatively simple decision matrices based on principal earthquake parameters (hypocentre, magnitude), or to try to constrain with observations a limited number of pre-designed tsunami scenarios. Both approaches involve significant uncertainty, which is usually not treated explicitly. For distant (e.g. trans-oceanic) tsunamis, there may be enough time to significantly better constrain the forecast with deep-sea tsunami observations (e.g. DART buoys[71]). This is not the case for local tsunamis, which impact the coast very rapidly. Moreover, deep-sea instrumental coverage is sparse in the world's oceans. eFlows4HPC will develop a Probabilistic Tsunami Forecasting (PTF) workflow that utilizes multiple FTRT numerical tsunami simulations using the multi-GPU software T-HySEA for seismic source ensembles. PTF is a new development[72] designed to deal with uncertainty estimation and is presently not yet optimized for efficiently utilizing resources on HPC or distributed architectures. To this end, eFlows4HPC will for the first time adopt dynamic workflows such as PyCOMPSs for operational PTF. This will allow much more efficient tsunami forecasting, including near-real-time source-estimation and novel post-processing tools enabling tsunami forecast and analyses in different computational environments from HPC centres to operational environments such as cloud solutions. In the context of **earthquakes**, producing maps of strong ground shaking after relevant earthquakes is based on the use of the USGS ShakeMap software[73]. ShakeMap generates maps that merge information coming from observed data (Intensity Measurements like peak ground acceleration and velocity, PGA and PGV, and spectral response, SA, at different periods) and from empirically derived ground motion models, GMM, that provide estimates of the ground motion as a function of magnitude and distance, and simplified proxies like Vs,30 to account for local site amplifications. These maps are well known for being a simplification when representing the ground motion accurately since the physics of the seismic wave field is not taken into account, except through empirical relations. It follows that the next step when producing more accurate maps of strong ground motion will consist of adopting full waveform simulations, which will take fully into account the three-dimensionality of the propagating media including topographic relief and soft sediments alluvial basins. In eFlows4HPC, given the inherent structural model uncertainties combined with the initial lack of detail of the earthquake rupture source, and the need to rapidly provide answers to the stakeholders (i.e., civil protection authorities and disaster risk managers) probabilistic rather than deterministic estimates of the experienced ground shaking, we plan to exploit the opportunities provided by the HPC resources through urgent computing for massive simulations using different input model parameters. We will also explore the possibility of using neural network analogy to simulations, specifically for sensitivity studies where a trade-off between accuracy and speed might be beneficial. By the same token, another aspect that will need exploration (and research development) is the *assimilation* of observed data into the simulations to achieve faithful reconstructions of the true ground shaking.

1.4.8 Innovation potential

Summarizing the aspects outlined in the previous sections, the eFlows4HPC project has multiple innovation potentials, coming from the specific workflow-oriented outputs and by the pillar specific ones. These innovation outputs have in some cases business exploitation potential while, in the others, may entail a high societal impact.

Project output	Innovation potential
eFlows4HPC workflow stack	Delivery of an innovative stack that enables the development of workflows that integrates HPC, HPDA and ML, enabling developers to reduce the time-to-solution and therefore reducing the time-to-market when applicable to industry problems. The stack includes optimized runtimes able to perform efficient data management and to reduce the time and energy required for the workflow execution.
HPCWaaS	Innovative concept that extends the FaaS model to provide a simple and transparent way of deploying workflows from repositories in federated HPC infrastructures. The innovation provided by the model consists on the reduction of the overheads of workflow deployment and execution, enabling a fast adoption of HPC by newcomers.
Model repositories	Provisioning of a model repository where trained ML models can be stored and later reused. This shortens the time-to-solution and the energy required to rebuild the models in the case of reuse. This approach simplifies the usage of ML by newcomers, increasing adoption.
Workflows for ROM in manufacturing	Innovative open source framework that integrate adaptive solution capabilities and ROM techniques within a single workflow. It focuses in the construction of Digital Twins contributes to Industry 4.0.
Workflows for Climate modelling	Holistic approach to end-to-end climate workflows that reduces workflow execution and computational cost, with transparent access to underlying infrastructures and open repositories to address reuse and reproducibility.
Workflows for Urgent computing for Natural Hazards	A new family of workflows for the modelling of earthquakes and associated tsunamis with urgency, able to deliver faster end-to-end runs, with more robust and reliable workflows.
Urgent computing policies	Proposal of access policies to HPC clusters that reflect on the urgent nature of natural hazard simulations, which radically differ from other resource access policies at public supercomputers.

Section 2. Impact

2.1 Expected impacts

Impact 1: *contribution to the realisation of the EuroHPC overall and specific objectives*
The project contributes to the following EuroHPC JU overall objectives: ***Impact 1.1:*** (d) to support an ambitious research and innovation agenda to develop and maintain in the Union a world-class **High Performance Computing ecosystem, exascale and beyond**, covering all **scientific and industrial** value chain segments, including low-power processor and **middleware technologies, algorithms and code design, applications and systems, services and engineering**, interconnections, **know-how and skills, for the next generation supercomputing era**;
The project will contribute with a European workflow platform (based on European components and libraries) focusing on the integration of HPC, HPDA, and AI; and with specific workflow solutions to application areas such as manufacturing, climate, and natural hazards, with scientific, social and industrial impact. Therefore, the project will contribute significantly to the software ecosystem for HPC from a European perspective. The text of the call is highlighted in bold for those aspects where the project is contributing.
Impact 1.2: (e) to promote the **uptake and systematic** use of **research and innovation results** generated in the Union by users from science, industry, including SMEs, and the public sector.
The project proposes the development of HPCWaaS methodology and the deployment of this methodology with a European workflow registry with the objective of easing the uptake and adoption of the project results by new and existing users of a wide set of communities. Specific workshop and training events will be organized to

communicate these results (see WP7). The project will implement a model repository and a libraries catalogue which will facilitate sharing, discovery and reuse of computation outcomes, i.e., trained ML and AI models, and of existing European software. A special case of adoption is found in the EDDL/ECV libraries, developed in the framework of the H2020 ICT DeepHealth project, which will ensure uptake and use of European software in the sensitive and crowded field of Machine Learning. The workflows developed by the Pillars in the project will be key for uptake in focused areas of application such as manufacturing, climate and natural hazards, especially from the EXCELLERAT, ESiWACE and ChEESE CoEs communities.

And to the EuroHPC JU specific objectives:

Impact 1.3: (e) to provide access to High Performance Computing-based infrastructures and services to a wide range of users from the **research and scientific community, as well as the industry** including SMEs, and the public sector, for new and emerging **data and compute-intensive applications** and services;

The project will develop innovative workflows in three key scientific and industrial areas. In particular, the workflows themselves will be instances of generic workflow templates that can be tailored to implement other instances with different functionalities. These workflows and templates will be delivered to the scientific and industrial communities (including SMEs) through the HPCWaaS methodology to facilitate its take-up and adoption, with a special emphasis on newcomers who may find a barrier in the current HPC policies and operation mode. The project areas cover both areas of interest for the scientific community and industry, including the SME in the project (DtoK). One of the envisioned project's outputs is a model repository with deployment option, enabling deployment of ML and AI models trained within HPCWaaS to conduct forecasts and analysis on new data.

Impact 1.4: (f) to support the development in the Union of world-class exascale and post-exascale High Performance Computing technologies, including low-power micro-processor and related middleware technologies, and their integration into supercomputing systems through a **co-design approach**, as well as their **uptake in large-scale and emerging application fields**;

The project includes co-design activities that focus on the optimization of specific application kernels for the EPI processor (task 3.3), as well as for other accelerators (task 3.2). The optimized application kernels will be used in the final version of the Pillar workflows.

Impact 1.5: (h) to achieve excellence in High Performance Computing applications for world-class performance through **development and optimisation of codes and applications and other High Performance Computing-enabled large-scale and emerging lead-market applications in a co-design approach,** supporting **Centres of Excellence in High Performance Computing applications and large-scale High Performance Computing-enabled pilot demonstrators and test-beds** for big data applications and services in a wide range of scientific and industrial areas;

The project contributes to this objective through the development of a new class of workflows, following a co-design approach between experts from different areas and involving the relevant CoEs and communities for the different target areas. The project will deploy its solutions in multiple EuroHPC pilot demonstrators, including pre-exascale EuroHPC systems (such as the MareNostrum5 and JUWELS) and the large MANGO heterogeneous prototype which will be used as a test-bed for parts of the workflow. Furthermore, the project has clear links with, and builds upon the success of, the CoEs EXCELLERAT, ESiWACE-2 and ChEESE.

Impact 1.6: (k) to **improve understanding of High Performance Computing** and contribute to **reducing skills gaps** in the Union related to High Performance Computing through awareness, **training and dissemination** of know-how;

The project's goals include the development of the eFlows4HPC software stack and the HPCWaaS methodology. Both aim at making it easier the development and usage of workflows that combine HPC, HPDA and AI. The project activities include the organization of workshops and training schools that will transfer the project methodologies and results to the relevant CoEs and industrial communities, contributing to the reduction of skill gaps in Europe related to HPC.

Impact 2: demonstrated relevance of the main target sector for European industry or society and in ensuring European technological autonomy in this field and in the Digital Single Market.

For the manufacturing industry, the pervasion of sensors together with industrial IoT enables completely new product functionalities. More and more features are realized by software. An essential part of the software is based on fast and reliable models that predict the behaviour of the physical environment. For the manufacturer of the physical asset this offers a unique opportunity, as nobody else than her/him can describe (= create a model for) the behaviour of the product. However, these physics-based models are usually too large to use during the

operation. Leveraging the technologies developed in pillar I the manufacturer can create a fast and compact Digital Twin in parallel to the physical asset allowing to create the above-mentioned novel product functionalities.
In the last decade, our understanding of climate change has increased, as society's needs grow for advice and policy. However, whilst there is a general consensus that climate change is an on-going phenomenon, there remain uncertainties, for example, on the levels of greenhouse gas emissions and aerosols likely that are emitted, or perhaps even more significantly, on the degree of warming and the likely impacts. Increasing the capability, comprehensiveness and performance of 'whole Earth System Models (ESMs) workflow', in order to represent with ever-increasing realism and detail new scenarios for our future climate, is the only way to reduce these uncertainties.
Natural hazards are a major concern for the EU and neighbouring countries. Several areas have been affected by destructive activity in the past, such as Greece, Italy, and Turkey for earthquakes and the whole Mediterranean coastline for tsunamis. A better assessment of risk and improved means for remediation are key to a resilient European society. Europe is also a major player in the reinsurance business, a key component of financial risk mitigation for high-loss events such as earthquakes and tsunamis. Companies such as Swiss Re, Munich Re or Hannover Rück are in the world's top-10 reinsurance companies and hence benefit from better assessments on the impact of disasters. As software-based solutions to trans-national disasters, the workflows developed with European technology are likely to benefit from the three Pillars of the Digital Single Market strategy proposed by the European Commission.
The project aims at providing a European-based solution for the development and deployment of workflows for both scientific and industrial applications. The fact that the solution does not focus on a single area or community, positions it for adoption by the European HPC community in general.
The project considers the optimization of specific kernels from the workflows in the project for the EPI, enabling a faster adoption and exploitation of such applications by European workflows.

Impact 3: Demonstrated innovation and productivity enhancement in the main target sector

The use of simulation based Digital Twins in manufacturing is currently limited mostly to linear cases, as the computational demand is very high for nonlinear problems. Reduced order models (ROM) are one possibility to shift the computational effort into an upstream offline phase. However, the effort still remains high such that HPC is a "must". Applying the workflows and algorithms developed in pillar I will boost the efficiency and speed of ROM creation thus allowing the creation of physics-based Digital Twins for highly nonlinear cases and opening up a broader field of industrial applications.
Climate workflows developed in pillar II will improve the productivity of the current simulations doing a smart usage and management of the computational resources needed to perform climate experiments. Dynamic workflows will be able to adapt simulations at runtime, by performing smart (AI-driven) pruning of members of a simulation and, therefore, avoiding unnecessary usage of resources. AI-based solutions for data analysis will bring innovative methods to extract knowledge from data produced by climate simulations leading to a better understanding of the climate system.
The project will be a key enabling component for a novel HPC access policy: urgent computing for civil protection. The concept, applied to earthquake and tsunami urgent simulations can be a breakthrough in the way in which HPC addresses societal challenges, effectively igniting the usage of public supercomputing resources in the field of disaster mitigation. By prototyping an urgent computing system for Tier-0 systems eFlows4HPC will pioneer the usage of HPC for immediate response to disasters, which might ultimately result in better and more accurate management of relief efforts.
The project aims at developing tools and methodology to increase the productivity on workflow development. In more detail, the HPCWaaS methodology aims at making easier deployment and usage of workflows, increasing the productivity of HPC users. Together with the establishment of a workflow registry and model repository, the project results will increase productivity in the selected sectors and beyond.

Impact 4: Effective integration of HPC technologies in the main target sector with measurable end-user metrics such as accessibility, scalability, performance, energy efficiency, reliability, and cost.

The proposed software stack aims at improving the efficiency of the HPC infrastructures, reducing the computational resources needed for each execution, therefore decreasing the total energy and economic cost. This will be demonstrated with use cases of application areas such as manufacturing, climate and natural hazards. Furthermore, the project includes tasks that aim at optimizing and tailoring specific project application kernels to heterogeneous HPC technologies (GPU, FPGA and EPI) and will also embed and adopt the EDDL/ECV libraries which support these technologies. With these activities, we guarantee the effectiveness of integrating

new HPC technologies. The project has specific tasks that will monitor the achievement of the indicated metrics by the project use cases.

Impact 5: Widening the use of and facilitating access to advanced HPC, big data and cloud infrastructures stimulating the emergence of the data economy in Europe.
The project will contribute to this with the proposal of the HPCWaaS methodology and its implementation, as well as by proposing workflow implementations to solve real and important scientific and industrial problems. The project intends to invest time in the standardization of execution environments with technologies like containers and virtualizations. This will contribute significantly to the vision of the European Open Science Cloud where resources can easily be reused. The different Pillars contribute to target areas where data is a fundamental asset. The project will implement in its processes, methodologies to ensure the secure sharing of personal data as well as proprietary, commercial and industrial data with the goal of ensuring its contribution to the data economy in Europe. With respect to not sensible data, the provision of a Data Catalogue and the Data Logistics Service will help to improve the visibility of the data used and produced by the participating Pillars. In particular, in natural hazards area, HPC is yet to be a main technological contributor for civil protection or the insurance companies. Nevertheless, the development of efficient, user-oriented simulation technology aims at enhancing the amount and quantity of information available quickly after an earthquake or tsunami. Effectively, and indirectly, resilience management stakeholders, and not only academic groups, will be able to integrate HPC technology in their policies and post-disaster protocols.

Expected impact KPIs	**Target value**	**Relevant to impact #**
Deployments of eFlows4HPC platform in supercomputing centres	≥ 2	1.1,1.3, 5
Number of complex workflows/models delivered with the eFlows4HPC	≥ 5	1.2,1.3, 1.5, 5
Number of newcomers evaluating HPC technology through HPCWaaS deployments	≥ 12	1.2, 1.3, 1.6, 5
Number of newcomers from SMEs evaluating HPC technology through HPCWaaS deployments	≥ 3	1.2, 1.3, 1.6, 5
Number of CoEs and user communities using developed workflows	≥ 3	1.3,1.5
Number of CoEs and user communities contributing with new workflows	≥ 2	1.3,1.5
Number of kernels optimized for accelerators (GPU, FPGA, and EPI)	≥ 6	1.4, 1.5, 2, 4
Number of workshops and trainings to transfer project results to CoEs and communities	≥ 4	1.6, 5
Improvement on development productivity of workflows that combine HPC, HPDA and AI. Lines of Code (LoC) and Cyclomatic Complexity (CC).	≥ 25%	3, 4
Complexity reduction in the workflow deployment in HPC systems (number of steps being automated)	≥ 80%	3, 4
Open European libraries and components in the eFlows4HPC	≥ 15	1.1, 2

platform		
Number of metrics defined and evaluated with the project Pillars	≥ 10	4
Number of data sources included in the Data Catalogue	≥ 6	5
Time to compute a Reduced Order Model starting from a Full Order Model and a predefined training campaign	< 12 h	2, 3, 4

2.1.1. Strategic Impact for EuroHPC

The project goals will contribute to reduce the technology dependence on US, by providing a European software stack for workflows and associated services, with optimization for the European processors (EPI).
In this project, we are framing the requirements, policies and impact of urgent access to Tier-0 HPC systems within the EuroHPC JU framework. By launching simulations of actual natural disasters, as opposed to simulating hundreds or thousands of hypothetical events for hazard estimation, we aim at offsetting the losses caused by interfering in their queue-based access protocols (e.g. shutdown of regular runs which might have limited checkpointing capabilities).

2.1.2 Societal impact

HPC is recognized as crucial in addressing grand societal challenges. The new workflows developed in the project will lead to more optimal use of computational resources, therefore reducing carbon and energy footprints of use cases defined in the Pillars. For example, in Pillar II, faster time to solution and HPDA/ML will enable more high-quality climate related research, potentially even for the scientists who had not use HPC previously (through HPCWaaS). This will improve the understanding of the climate system, and lead to improved mitigation and adaptation strategies for future climate change.
The work done in Pillar II will contribute to the achievement of the United Nations Sustainable Development Goal #13 ("Take urgent action to combat climate change and its impacts") by improving current climate simulations and, therefore, increasing our current knowledge of the climate ecosystem and delivering better data to end users to develop new and better approaches to climate adaptation.
Disaster resilience (UNISDR, 2005) is determined by the degree to which individuals, communities, and public and private organisations are capable of organising themselves to learn from past disasters and reduce their risks to future ones, at international, national, regional, and local levels. Our activities in the field of natural hazards, particularly tsunamis and earthquakes, are directly related to improving our society's resilience to such events. Simulations will provide us with unprecedented accuracy when analysing the impact of such events, which is a key component to remediate its associated damage and thus reduce the risk of suffering losses related to inefficient mitigation efforts.
Efforts towards model registry to share outputs of computations previously done and to reuse workflows in the repository with HPCWaaS will significantly increase transparency of social-relevant scientific findings, enabling their repeatability, reproducibility and verification even by untrained civil scientists.

2.1.3 Industrial impact

The project potential impact is to enable the integration of data-intensive simulation in the lifetime management of industrial products, thus achieving the "Industry 4.0" vision. Digital Twin-based solutions rely on physical effects best known to the manufacturer of the physical asset. Therefore, all the technologies that enable the manufacturer to create a fast, powerful and comprehensive Digital Twin are highly relevant to the manufacturing industry. In particular, the reduction of already existing (during the product design and verification) physics-based models have a high potential and put the manufacturer in a leading position, as he can use of already existing product knowledge. This will be achieved by the use of an open source, **European**, set of technologies, thus fostering European leadership in the sector.
Furthermore, the project will put in contact leading Academic institutions in the sector with an industry giant (Siemens) providing a unique opportunity for the simulation-based Digital Twins and for the entire workflow to be integrated in the pipeline of one of the world giants in the sector of advanced manufacturing.
From a business and economic point of view, the outcomes from the project will help the industrial partners to increase their software portfolio with new components, such as high-level workflow management, AI model management, to more efficiently assist professionals in their digital strategy (diagnostic, decision-making processes, automation…). This project will also provide the opportunity to develop new solutions in the big data analytics / AI

market while **enriching the industrial partners' know-hows**. Supercomputing systems and technology providers will benefit from the democratization brought by the project regarding the usage of HPC systems for analytics & AI applications.
The industrial impact of the development of the three application Pillars through the integration of HPC and data analysis techniques in a workflow-based environment will also improve the competitiveness of European software **small and medium enterprises** needing to implement advanced solutions and software products related to Big Data management and processing. The project industrial impact is relevant for European players not only in the manufacturing area but also in the wide digital market, as they can benefit from the new software tools and HPC techniques that will result from the project activities. The project results will also represent a set of innovative technologies that serve for software companies working in the area of data science applications.

2.1.4 Standardization Activities

The project will take into account different standards and standardization initiatives for its activities. With regard to standards for workflows, the partners will consider how to contribute and conform to TOSCA. The overall workflow components in eFlows4HPC will be described with an Extended TOSCA syntax, which feed the Ystia orchestrator to manage the applications' lifecycle on hybrid infrastructures. Atos plans to push some new components to the TOSCA standard.
The PyCOMPSs programming model will provide the logic of the different components of the overall workflow. While the PyCOMPSs programming model can be considered as the task-based OpenMP version for Python, such standard is not considered yet in Python. However, this has been discussed in fora such as the SC19 Python for HPC BoF. If the opportunity of defining such standard for Python arises, BSC aims to promoting it following its tradition of contributing to the ARB OpenMP standard.
In the development of the eFlow4HPC core components, services and models as well as domain-specific workflows and use cases, several interoperable aspects will be taken into account such as the use of (i) standard data formats for data exchange (e.g. HDF5, NetCDF, JSON), (ii) service interfaces (e.g. OGC-WPS/WCS for server-side processing), and (iii) well-established (de-facto standard) community-specific conventions (e.g. CF, Climate Forecast for climate data).
The proposal will follow initiatives that define such standards, for example, following the SC BoF on HDF5 where BSC is making proposals for its Virtual Object Layer (HDF5 VOL). Similar activities will be performed when considered appropriate.
When defining the HPCWaaS methodology, the partners will consider which standards are relevant and available to ensure that what will be delivered is compliant to given standards to further address reuse. Examples are REST interfaces for the service orientation approach and container standards for packaging, among others.
With regard to some of the community specific libraries and data, they can constitute a solid basis for further cross-community standardization effort for instance under the umbrella of the Research Data Alliance (RDA).
BSC and UPV are partners of the OpenPower Foundation, which aims to enable data centres to rethink their approach to technology. Member companies are enabled to customize POWER CPU processors and system platforms for optimization and innovation according to their business needs. These innovations include custom systems for large or warehouse scale data centres, workload acceleration through GPU, FPGA or advanced I/O, platform optimization for SW appliances, or advanced hardware technology exploitation. Therefore, the eFlows4HPC project has direct links with the Foundation and will match the output of the project with all suitable specifications and standardization efforts within OpenPower.

2.1.5 Contribution to European Policies, Regulations and Norms.

Pillar III aims at drafting an access policy for urgent computing at HPC infrastructures that will be made available to the EuroHPC JU. A protocol for urgent access to Tier-0 and Tier-1 systems will be developed that will include automation rules, an evaluation of the cost and benefit of different scenarios, and making recommendations for the governing bodies of such systems regarding its implementation. This should include not only availability and QoS of the infrastructure but also early risk estimates related to the hazard, which should be standardized across different hazard types.

2.1.6 Barriers to Achieving the Aspired Impacts

Besides the strategic impacts outlined by the call, eFlows4HPC aspires to bring significant impacts, which are achievable either within the project lifecycle or after full-scale adoption of the project outcomes by a wide range of stakeholders. The consortium is fully committed to mobilise all resources required in order to realise the envisaged impact while dissemination, stakeholder engagement and communication activities (as described in section 2.2) are expected to be instrumental for the multiplication of the envisaged eFlows4HPC impacts. Nevertheless, although the

project impacts are tangible, there always exist certain external and internal factors and barriers over which the project consortium has limited control and that may affect and possibly hinder the impact achievement. Such barriers are broader than the scientific, business, and technical risks identified in section 3, and shall be carefully considered and appropriately tackled, since they may significantly affect a core success factor like user acceptance.

Staffing difficulties: The demand for IT personnel is quite high, and staffing is sometimes an issue in EC funded projects. This situation also may differ depending on the global situation of the economy. However, the consortium brings together stable and committed teams that guarantee the workforce of the project.

Barriers for adoption: Even though the developer workflow syntax and the end-user interface will be designed to ease the use and promote the adoption of the project tools; the entry barrier to new environments always requires some collaborative effort. Communities tend to adopt those tools created and available in their environment. That is the reason it is so important to work in collaboration with the user communities in the design of the syntax and interfaces, following a co-design approach between technology providers, workflow developers, and end-users. Additionally, training and user community workshops should help on to reduce this barrier.
Another barrier for adoption is the existence of previous *competitors*. In this case, we have selected communities with no predominant solution being used or participants on the Pillars willing to explore the project proposed solutions. Through the collaborative effort in the project, and through the validation of the use-cases in the Pillars, we aim at demonstrating the advantages of our approach and achieve community adoption.

A more detailed analysis of the project can be found in the form of a SWOT table in Annex E.

2.2 Measures to Maximise Impact

2.2.1. Dissemination and Communication strategy (incl. Data Management Plan)

Effective communication is essential to ensure that the project reaches its intended beneficiaries and to promote the value of research and industrial ideas to key stakeholders and to the wider public. In order to achieve this, it is necessary to define a clear communication strategy at the outset of the project, identifying target audiences and selecting the appropriate means of communicating with each of them (see table below). The Consortium benefits from communication channels of well-established partners and a wide range of contacts in industry and research, upon which it will draw its strategy in order to communicate project benefits.

Target audience	Key messages	Value to target audience
HPC users of manufacturing, natural hazards, climate	eFlows4HPC facilitates the deployment and usage of complex workflows in federated HPC environments.	Engagement of newcomers to HPC by increasing the productivity and reducing the learning curve, and applicable to other potential HPC sectors beyond the ones in the proposal. Bridges the gap between end users and HPC experts.
Application/workflow developers who want to combine HPC, HPDA and AI	1) Ease the development of complex workflows that include HPC, HPDA and AI components; 2) Access to an execution platform that automatically parallelize the workflows.	Increase on application development productivity in distributed computing systems. Integrated platform for HPC, HPDA and AI. Ease of deployment and usage.
HPC and related research community, with special emphasis on HPC CoEs	Availability of workflow methodologies and specific application workflows that can be reused and easily tailored to other potential use-cases.	Potential new users of the HPC CoEs' applications engaged by availability of new workflows ready to use. Support to roadmap definition for policy makers. Methodologies that can be exploited towards HPC CoEs sustainability. Increased recognition of workflow technologies.
Industrial stakeholders	Provide real-time methodologies in manufacturing design by using digital twins that reduce costs and speed-up	Improved competitiveness for European companies and SMEs through access to ready-to-use applications.

	production. Increase competitiveness of European HPC vendors.	Availability of new European software platforms.
Policy makers	Provide techniques that reduce power consumption. Proposal of protocols for urgent computing in natural hazards using HPC. Building upon European expertise in HPC. Promotion of European software, expertise and know-how. Provide tools to federated HPC platforms for complex and critical workflow executions.	Opening up opportunities to enhance European competitiveness in burgeoning new markets. Investment on innovation increases competitiveness and progress of society.
Society and general public	Reduction on time to react on a natural hazard (for tsunamis and earthquakes). Reduction of climate modelling time Improvement on manufacturing processes.	Improvement of hazard mitigation protocols thanks to European HPC. Contribution to the processes that evaluate climate change impact. Reduction of acquisition price of manufactured products

Communication activities will seek to:

1. Raise awareness about the project and its results: key goal of the project in building European expertise in the development of complex workflows that combine HPC, HPDA and AI, with an emphasis in the areas of manufacturing, climate and urgent computing for natural hazards.
2. Update key stakeholders on project progress.
3. Build community around the project

To do so, several communication actions will be performed:

- A distinctive **brand identity** will be created for the project. This will be applied consistently across all project dissemination materials and templates in order to reinforce awareness of the project.
- The project **website** will form the central hub of the communication and dissemination activities: it will bring together papers published in scientific journals and information about events where the project will be presented, as well as allow visitors to download dissemination material such as the project flyer, posters, videos, whitepapers, the brand centre, etc. Related communication activities such as press releases, brochures and posters will be put online and will be distributed to the appropriate target groups (media organizations, scientific community, industry, policy makers and civil society).
- A presence on **social media** helps foster a digital community and complements face-to-face networking activities. In eFlows4HPC, we envision opening and updating a Twitter profile and a LinkedIn group. The message will be adapted to each one of the channels. For instance, Twitter will be used in a wider range of communications, from the plainer information to the one targeting industry or research community. Twitter allows for rapid interaction with the public and stakeholders. In the LinkedIn group, the information posted will be meant for professionals and it will be enriched with more in-depth technical details. In LinkedIn the project results will also be shared in well-established groups on the same field.
- With the aim of disseminating the eFlows4HPCsoftware architecture and results to European industry outside the project, the social media strategy will be complemented with the production of **leaflets and factsheets**, a useful and attractive way of providing the key facts about the project at a glance, easily distributed at events. Press releases can be prepared and released as required to disseminate project results to either the specialist in the HPC domain or the general press in different languages.
- Understanding that our society is increasingly consuming information by visual means, the dissemination team will produce **short videos** during the project, in English with subtitles in local languages.
- The consortium will present the results on top **tier conferences via peer-reviewed papers, workshops and keynotes**. Below can be find the main conferences or exhibitions –where partners have long success records and, in many cases, relevant presence in selection committees, as chairs, etc. Three **joint community workshops** with HPC CoEs and eFlows4HPC will also be considered (see task 7.2 in WP7).

- A total of **three training courses** (one per pillar) to engage with the related communities related to the Pillars as well as a **hackathon** across Pillars will be organized. A detailed training plan will further explain the training activities (D7.1). In addition, training activities addressed to students will also be considered by all partners to contribute; for example, the SISSA mathLab team will organize a thematic summer school that will help to spread the developed methodologies to a broad audience of users and developers.

The following journals and conferences will be taken into account in the Dissemination and Communication plan where further details will be given:

- Journals: Future Generation Computing Systems, Journal of Grid Computing, International Journal of High Performance Computing Applications, Concurrency and Computation: Practice and Experience, IEEE Transactions on Parallel and Distributed Systems, Journal of Parallel and Distributed Computing, Journal of Supercomputing, IEEE Internet Computing, Geoscientific Model Development, Computer Methods in Applied Mechanics and Engineering, Computational Mechanics, International Journal for Numerical Methods in Engineering, Journal of Computational Physics, Journal of Advances in modelling Earth Systems, Solid Earth, Computational Geosciences, Computers and Geosciences.
- Conferences and (industrial) events: SC series (Supercomputing), ACM International Conference on Supercomputing, IEEE Cluster, Euro-Par, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Euromicro International Conference on Parallel, Distributed and Network-Based Computing, Int. Conference on High Performance Computing and Simulation, ACM High-Performance Parallel and Distributed Computing, IEEE International Parallel and Distributed Processing Symposium, Computational Science - ICCS, IEEE Big Data Conference, European Geosciences Union General Assembly, American Geosciences Union meeting, World Congress in Computational Mechanics, European Congress in Computational Mechanics, Coupled Problems, Platform for Advanced Scientific Computing (PASC),
- Other events where the partners plan to participate and disseminate the project results, although in most cases there is no associated publication are workshops or BoF sessions at ISC and SC conferences, BDEC events, the EuroHPC Summit Week conference series, FocusCoE events and RDA conferences.

For further control dissemination and communication strategies, table below shows an initial set of key performance indicators that eFlows4HPC will establish to track dissemination activities and their effectiveness over the course of the project (adjustments will be made if necessary).

Dissemination channel	**KPI**	**Measure**
Scientific Publications	Papers published, both in scientific venues and Journals	At least 15 publications in total
Academic and Industrial Events	Events organised, including conference booths, tutorials and workshops (with significant attendance, i.e. above 25 people)	At least 6 events with attendance and a booth in an industrial-related event
Website	Visitor statistics (number of unique website visitors and their location captured by Google Analytics)	Increase in at least 1% the returning visitors each year
Dissemination material	Number of posters Number of project videos	At least four posters At least three short videos
Training courses	github repository Training event page Number of attendees	At least one per pillar One hackathon across Pillars One summer school, SISSA
Joint community workshops	Organization of workshops for the Pillars	Three joint workshops to engage with the related community of each pillar

Open Access

The consortium is committed to provide at least green open access wherever feasible following the provisions of Horizon 2020 guidelines. The project will guarantee open access to scientific peer reviewed publications by depositing a machine-readable electronic copy of the publisher's final version of the paper or a final peer-reviewed manuscript accepted for publication, respecting always the embargo period. Each publication will be accompanied by bibliographic information, publication date, metadata about project funding (name of the action, acronym and grant number), date of release in open access, and a persistent identifier. For the coordinating partner BSC, the institutional repository UPCommons[xxv] will be used. For the rest of the partners' publications, the coordinator will remind them about the H2020 open access publication policy by informing them about the available repositories such as Zenodo provided by OpenAire. In any case, both options enable third parties to access, exploit, reproduce and disseminate at no cost.

With regards to open source software, that is claimed to be the result of a project, it will also be available from the project page or open repository including a readme file with the acknowledgement sentence.

2.2.2 Plan for exploitation of results

Exploitation stakeholders

The exploitation plan is drafted as a collaboration of the project partners, which include supercomputing centres (BSC, FZJ, PSNC), vendors and industrial partners (Atos, SIEMENS, DtoK Lab), solid state and climate institutes (CMCC, NGI, INGV, AWI), research centres (CIMNE, INRIA, SISSA), and Universities (ETH Zurich, UPV, UMA).

Exploitation objectives

The preliminary exploitation plan for eFlows4HPC is drafted based on different axes according to the different solutions generated by the project and on the different roles played by the partners participating in the proposal.

A first axis is based on the **eFlows4HPC platform,** composed of the **eFlows4HPC software stack** and the **HPCWaaS concept**, which together offer a unique opportunity to define a European platform for the development and deployment of workflows that combine HPC, HPDA and AI in supercomputing infrastructures, including EuroHPC pre-exascale infrastructures. With this objective, the eFlows4HPC software stack and HPCWaaS methodology will be deployed and demonstrated in production in the infrastructures available to the project (PRACE Tier-0 and Tier-1 supercomputers). By demonstrating these technologies in the premises accessible to the project, the partners as a whole aim at promoting this solution for all EuroHPC infrastructures. This exploitation will be done initially with the workflows developed in the project, but others will be added according to the centres' and communities' needs.

A specific case is found in the **industrial partners** (Atos and DtoK lab) that plan to productize the developments of **the orchestration and AI engineering** and to integrate them into their commercial products and software suite and to use them in application domains where the analysis of **Big Data through workflow** frameworks and libraries represent a key solution both for performance and functionality.

Another axis is based on the different workflows developed by the three Pillars. Each of these workflows will be exploited by the communities in a slightly different manner.

The partners involved in Pillar I aim to take advantage of the HPCWaaS methodology to explore **commercial usage** mechanisms of the workflows developed in the Pillar. Given the capability of ROM to enable real-time accurate simulations, this project will be an excellent benchmark to demonstrate the impact of such models on applications of **practical industrial interest** related both to the design phase and to digital twins where complex computational mechanics phenomena are involved. Furthermore, thanks to the interaction with experts in the topic, the exploration of the use of AI techniques in the context of adaptive remeshing together with solvers is a great opportunity to enhance the capabilities and potential impact of the software. At the same time, this creates new scientific perspectives and possibilities for collaboration in this community, which is rapidly impacting many independent research sectors including computational mechanics. This is reinforced by the presence of the **industrial partner** in the Pillar, since a major research line of Siemens Corporate Technology is the **investigation of Digital Twins** in the operation phase. The integration of the multiphysics solver Kratos in the eFlows4HPC software stack provides an invaluable opportunity to derive small yet accurate reduced order models (ROM) in an offline phase, that can run in the future even on enhanced state of the art control hardware like SIMATIC-S7 during operation, thus **influencing directly the further developments of Siemens control hardware**.

Concerning the climate pillar, BSC and CMCC are already running simulations in the context of the CMIP experiments and disseminating them through a data-node hosted at BSC. The improved workflow capabilities

[xxv] http://upcommons.upc.edu/handle/2117/23714

provided by eFlows4HPC developed in the project will contribute to **enhance climate simulations** workflows running both at BSC and CMCC in terms of time-to-solution, carbon footprint, and overall end-to-end simulation management, thus providing a more integrated and holistic approach joining into the same end-to-end workflow numerical models, HPDA and ML components. The workflows developed by Pillar II will be **further exploited by the EsiWACE2 CoE**, enabling the community to run climate simulations in the upcoming pre-exascale machines.
Pillar III addresses urgent computing for civil protection. The concept, applied to earthquake and tsunami urgent simulations, can be a breakthrough in the way HPC addresses societal challenges, effectively igniting the usage of public supercomputing resources in the field of disaster mitigation. The workflows tackled by the pillar focus on end-to-end urgent computing of earthquakes and tsunamis. The new approaches to general workflows within natural hazard applications developed in this project may demonstrate applicability to other project areas in the broader multi-hazard and multi-risk domain in the future, improving capabilities and catalysing future innovation in neighbouring applications not covered in this project, with the ultimate goal of **disaster risk reduction**. The workflows developed by Pillar III (UCIS4EQ, PTF) could be **transferred to the ChEESE CoE** and **further exploited by the solid-earth community** in the frame of this CoE, such as the research and consulting activities performed by the Norwegian Geotechnical Institute (NGI) or the Instituto Nazionale di Geofisica e Vulcanologia (INGV). The added value of this Pillar relies also in the creation of the **policy tender proposal for the Urgent Computing** deployment. By prototyping **Urgent Computing for EuroHPC Tier-0** systems in the novel way, the Pillar partners will pioneer the usage of HPC for immediate response to disasters, which might ultimately result in better and more accurate management of relief efforts.
Another axis comes from the exploitation of machine learning libraries. The project will also extend and optimize some **European machine learning libraries** such as HeAT, EDDL/ECV and dislib, which will be further **optimized and tailored to heterogeneous resources such as FPGAs and EPI**. While some of these libraries are already used in production in some of the supercomputing centres in the consortia, their extensions and enhancements in the project will increase their competitiveness.

Exploitation strategy

The exploitation activities task will begin with the analysis and exploration of relevant eFlows4HPC applicable markets, which will be turned into an **in-depth market analysis**. This analysis will be presenting the competitive landscape with regard to solutions coming out of the three Pillars' use cases and for the main technological components, i.e. software stack, workflow services, etc. This will allow carving a niche positioning for the eFlows4HPC outcomes and build the **value proposition** for the adoption of the project's **key exploitable assets** by the community.
The next step will be to define a comprehensive **exploitation strategy**, laying out exploitation scenarios for the sustainable uptake of the different components and an action plan to realize this. In addition, the strategy will incorporate the individual exploitation perspectives of the different partners. Atos will lead the exploitation activities with the contribution of all industrial and research partners with different exploitation interests.

Individual early exploitation plans

***1. BSC**:* The contribution of BSC to eFlows4HPC comes from multiple aspects, as coordinator of the whole initiative, and for its contributions from the Computer sciences, Earth sciences, and CASE departments. The Workflows and Distributed Computing group has been developing PyCOMPSs/COMPSs, Hecuba and dataClay during the last years. PyCOMPSs/COMPSs has been used in production in the different versions of MareNostrum to implement different scientific workflows and applications, integrated with the Hecuba and dataClay data management solutions. BSC aims to deploy the whole eFlows4HPC software stack and the HPCWaaS methodology in MareNostrum 5, one of the three pre-exascale systems to be deployed in Europe in 2020 and to exploit it, among others, with the workflows developed in eFlows4HPC.
In the aspect of natural hazards, BSC has developed the UCIS4EQ workflow that enables end-to-end urgent computing of earthquakes. Within the ChEESE CoE (also coordinated by BSC) we have linked the workflow with state-of-the-art seismic modelling engines and thus have engaged with the earthquake simulation community in Europe. This community will benefit from the developments of eFlows4HPC in terms of accessibility and performance of the workflow. Furthermore, the CoE includes two primary end-users for UCIS4EQ, which are INGV (also a partner in this project) and the Meteorological Office in Iceland (en.vedur.is). Both institutions hold responsibilities in the monitoring of seismic activity in their respective countries and boast strong ties with the local civil protection. They have stressed interest in adopting our technology in their own services. By means of the project ARISTOTLE, coordinated by INGV, we will also have access to the coordinated European civil protection ecosystem. We aim at establishing UCIS4EQ as an operational service for civil protection by the end of the project. UCIS4EQ has a modular structure that can link with different HPC codes and has mechanisms to

steer simulations by adding/subtracting events to the simulation pool in order to deal with circumstances such as updated information on the events, additional data not present at the onset of the urgent run or further manual requirements. By means of the USER project, and also within the ChEESE CoE, we are exploring ways of potential commercial use of UCIS4EQ in the sectors of reinsurance, where early assessments of earthquake affectation can be financially crucial, and also in risk analysis.
Concerning the climate pillar, BSC is running simulations in the context of the CMIP experiments and disseminating them through a data-node hosted in the institution. The improved workflow developed in the project, will contribute to deliver better climate simulations reducing the computational resources and, therefore, reducing the carbon footprint of such experiments. This workflow will also be disseminated within the ESiWACE2 CoE (where BSC co-leads the WP about "production runs at unprecedented resolution on pre-exascale supercomputers"), contributing to run climate simulations in the upcoming pre-exascale machines.

2. CIMNE contributes to the project with the open source framework Kratos. The integration of such solver in the eFflows4HPC workflow provides an invaluable opportunity to spread the usage of Kratos. Furthermore, the integration of the software with AI solvers and the interaction with experts in the topic will enhance the impact of the software and open new possibilities for collaboration with leading experts in the sector, an extremely appealing possibility since AI is one of the most rapidly growing communities as of now.
Furthermore, CIMNE plans to take advantage of the HPCWaaS mechanism to explore, jointly with the pillar partners, commercial usage mechanisms. Such mechanisms will be designed preserving the openness of the framework (essential for collaboration with academia and industry) and will focus on providing a premium ease of use of the integrated simulation and model-order reduction package.

3. FZJ is the world-leader in the field of modular supercomputing. It follows a strategy to extend HPC resources with specialized modules like boosters, machine-learning clusters, and quantum computing elements. Such heterogeneous environments can only be exploited to the fullest extent if the users are able to compose large computing workflows with parts that can be placed on these modules. Therefore, the flexible workflow platform proposed in the project is absolutely crucial for this partner.
Juelich Data Logistics Service is currently used by researchers from FZJ to collect air quality measurements from stations spread around the world. Further extensions of this service planned in the project, e.g., towards data streaming and improvements in the ease of use will benefit the climate modelling community and can widen the usage of the service to support the process of large sets of scientific data gravitating towards scalable processing facilities.
Scalable machine learning library HeAT, which is developed by researchers from FZJ (among others) aims at the efficient use of heterogeneous resources by abstracting high performance computing libraries as well as GPU-specific ones. The opportunity given by the eFlows4HPC to include and optimize even more special hardware classes (FPGA, EPI) is critical for HeAT to keep its advantage over existing ML libraries.
In recent years, we have observed increased interest in containerization technologies among the users of the Juelich supercomputer systems. The planned establishment of set of best practices in this field along with a model and workflow registry will underpin the long-time effort and commitment of Juelich Supercomputing Center to Open Science.

4. UPV will perform an academic-oriented exploitation plan including dissemination through external seminars and tutorials and master courses taught at UPV premises. On a different venue, UPV will analyse industrial oriented exploitation of the results through the creation of spin-off companies, the transfer of technology to existing UPV spin-offs, and the exploitation through new research projects.
UPV will also contribute to the open-source availability of the tools and platforms developed in eFlows4HPC that can be used by industrial and academic users. In particular, UPV will exploit the expected results (e.g., machine learning techniques developed in collaboration with the application Pillars) in other collaborative projects, transferring this output to potential end-users of the UPV and Valencia region network, and/or through the bi-lateral contracts with interested companies.
One important exploitation plan for UPV is the adoption of the MANGO prototype for HPC applications and the demonstration of energy-efficient computation of new heterogeneous architectures. UPV will promote the eFlows4HPC results in the scope of the prototype.

5. Atos will leverage the eFlows4HPC project on the HPC, data analytics and Artificial Intelligence (AI) technology convergence to enable a scientific workflow platform. The project will provide new insights of key interest for Atos's customers. Atos plans to combine the results of the eFlows4HPC project with those of Mont-Blanc (3 & 2020), LEXIS, CloudDBAppliance and SAGE/SAGE2 projects, in addition to its own developments on the BullSequana S & BullSequana X product lines to improve the delivery of global HPC/Big Data/AI solutions for the 2021/2024 timeframe. With this effort, Atos will be able to provide application workflow support as well as AI models management support in line with best practices and key European scientific users requirements.

Regarding the orchestration and AI engineering, Atos intends to productize the developments, and to integrate them into its commercial products & software suite. Some elements will be pushed into the TOSCA standard. Data centres and academics may then benefit from a mature/productized flexible orchestration and AI engineering solution.

6. DtoK Lab: The project results related to the availability of workflow tools and methodologies are of interest for DtoK Lab that will exploit them in the data analysis projects and consultancy activities it is carrying out with customers in different application domains where the analysis of Big Data through workflow frameworks and libraries represent a key solution both for performance and functionality. DtoK Lab will contribute to the project with high-level mechanisms for HPC data analysis such as ParSoDA that will be integrated in the eFlows4HPC stack. At the same time, the DtoK Lab team will exploit the solutions of the project for enhancing and extending its software portfolio and for improving the company portfolio in the area of service-based Cloud applications for end users in a scientific and business domain.

7. CMCC*:* in the climate change community, is contributing to the CMIP experiments both in terms of simulations (through CMCC models, like the CMCC-CM2) as well as data dissemination (through two ESGF data nodes hosted at the CMCC SuperComputing Centre). As a technology provider, CMCC develops Ophidia, an open source framework addressing HPDA for climate change. The improved workflow capabilities provided by eFlows4HPC will contribute to enhance climate simulations workflows running at CMCC in terms of time-to-solution, carbon footprint, and overall end-to-end simulation management, thus providing a more integrated and holistic approach joining into the same end-to-end workflow numerical models, HPDA and ML/DL components.
CMCC plans to deploy and exploit the whole eFlows4HPC software stack on its Zeus SuperComputing system - which is hosting the CMCC Data Science environment - to better support climate scientists with respect to ML/DL over HPC.
CMCC is actively contributing to HPDA and ML/DL research for climate change; eFlows4HPC will allow the consolidation and further improvement of on-going efforts in this field as well as of the data analytics operational workflows running at the centre to support scientists' needs.
The project results will be also disseminated within the ESiWACE2 CoE, where CMCC co-leads the WP about "Data post-processing, analytics and visualization" contributing to the development of large-scale, HPC-enabled data systems.

8. INRIA contributes to the project with the open source framework MMG. The integration of such solver in the eFlows4HPC workflow provides an invaluable opportunity in spreading the usage of this platform. Furthermore the exploration of the use of AI techniques in the context of adaptive remeshing, thanks to the interaction with experts in the topic, is a great opportunity to enhance the capabilities and potential impact of the software, at the same time creating new scientific perspectives and possibilities for collaboration in this community which is rapidly impacting many independent research sectors including computational mechanics.
Via the interaction with the pillar partners, we also hope to be able to showcase the economic interest of open-source software models, even in the context of applications requiring the integration of several packages (computational module, meshing/re-meshing/ reduced order models).

9. SISSA will contribute to the project through the development of reduced order models, which will be integrated within the eFflows4HPC workflow stack with the aim of taking full advantage of next generation HPC environments.
SISSA has already developed ITHACA, a collection of stand-alone libraries for model order reduction which are already freely available online, and the implementation of these reduced order techniques in the eFlows4HPC workflow will enhance their spreading and will open future possibilities for collaborations.
Moreover, given the capability of ROMs for real-time accurate simulations, this project will be an excellent benchmark in order to demonstrate the impact of such models on applications of practical industrial interest related both to the design phase and to digital twins where complex computational mechanics phenomena are involved.

10. PSNC will address a key enabling component for a novel HPC access policy: Urgent Computing for civil protection. The concept, applied to earthquake and tsunami urgent simulations can yield a breakthrough in the way in which HPC addresses societal challenges, effectively igniting the usage of public supercomputing resources in the field of disaster mitigation. The creation of the policy tender proposal for the Urgent Computing deployment for further use is predominant in the deployment of the real case. By prototyping Urgent Computing for EuroHPC Tier-0 systems in the novel way, PSNC with corresponding parties will pioneer the usage of HPC for immediate response to disasters, which might ultimately result in better and more accurate management of relief efforts.

11. UMA will contribute to the project with a tsunami numerical code for tsunami simulations: Tsunami-HySEA. The integration of this solver with the eFlows4HPC software stack will provide a great opportunity to improve the usage of the code among the tsunami community, from modellers to end-users. Moreover, the combination of the numerical model with AI techniques will improve its applicability in Probabilistic Tsunami Forecasting (PTF) and

Probabilistic Tsunami Hazard Assessment (PTHA). Furthermore, UMA plans to take advantage of the HPCWaaS mechanism to explore, jointly with the pillar partners, commercial usage mechanisms preserving the openness of the designed models.

12. INGV: To provide realistic and unbiased intensity forecasts in the aftermath of a geophysical event (e.g., earthquakes, tsunamis, volcanic eruptions) is one of the activities of INGV in support of decision making. The expected benefits for INGV include improved earthquake and tsunami forecasting capabilities for early warning and emergency computing purposes using workflows for HPC, in particular within our core applications for Probabilistic Tsunami Forecasting (PTF), as well as in the Urgent Computing Integrated Services for EarthQuake (UCIS4EQ) application. The new approaches to general workflows within natural hazard applications in eFlows4HPC (earthquakes and tsunamis) may demonstrate applicability to other project areas in the broader multi-hazard and multi-risk domain in the future, improving capabilities and catalysing future innovation in neighbouring applications not covered in this project, with the ultimate goal of disaster risk reduction.

13. AWI has long-term experience in climate research and participates in CMIP6 activities with its climate model. AWI contributes to the project with a coupled climate model OpenIFS-FESOM2. The ocean part of the model system - FESOM, is a next-generation ocean model, formulated on an unstructured mesh. Two main components of the climate model are fast and scale well. This allows performing ensemble simulations with relatively high spatial resolution, as well as single extremely high-resolution simulations. This type of research requires a new approach to the simulation workflows and data handling, and in eFlows4HPC, AWI will cooperate with the best European specialists in the field to implement it for OpenIFS/FESOM2. This is also a unique opportunity to modify AWI data processing workflows since challenges in the analysis of high-resolution data currently hinder scientific progress in the area of high resolution coupled climate modelling. Implementation of the OpenIFS/FESOM2 system in the workflows developed by eFlows4HPC will increase its visibility and make the system easier to use, which, in turn, will increase adoption as a tool among the climate modelling community. AWI will also greatly benefit from the opportunities to apply ML technologies to climate research/modelling for big data analysis, storage optimization and replacement of physical parameterizations in the climate model. eFlows4HPC will allow to interact with providers of ML solutions and provide climate-related test cases. The project will also help to identify performance bottlenecks and prepare OpenIFS/FESOM2 for exascale era.

14. ETH Zurich: The contribution of the Seismology & Wave Physics Group at ETH Zurich falls into 3 categories: (1) Structural models of anisotropic P- and S-wave velocities and density for selected regions of interest, including parts of Iceland, Northern Italy, the Sea of Marmara region, the Iberian Peninsula, and the Aegean. These will be ensemble models, i.e., suites of models that are consistent with the seismic data used to construct them. These models will be freely available for download and further use, e.g., in seismic hazard studies. (2) Raw ground motion simulation data produced with the previously suite of models will be equally made publicly available. (3) Probabilistic shake maps derived from the simulations will be made available in the form of data files and movies. The results of eFlows4HPC will be exploited by ETH Zurich by offering software to researchers outside the project able to produce their own ground motion metrics.

15. Siemens (Corporate Technology) supports the Business Units of Siemens with project related research. A major research line of Siemens Corporate Technology is the investigation of Digital Twins in the operation phase. Here the creation of fast and reliable executable models that can be used to run during runtime of physical objects is of special interest. Especially for nonlinear effects, the creation of ROM is currently very time consuming and laborious. A further challenge is the estimation of accuracy of possible operation alternatives. To give a reliable suggestion for operation alternatives many possible cases have to be evaluated and compared in real-time.
Both challenges require access to an efficient computing infrastructure as well as efficient workflows as those developed in eFlows4HPC. The integration of the multiphysics solver Kratos in the eFlows4HPC software stack provides an invaluable opportunity to derive fast and small reduced order models (ROM) in an offline phase, that can run in the future even on enhanced state of the art control hardware. Siemens expects to include ROM based on the eFlows4HPC technology in Systems like SIMATIC-S7 to enable model-based control, thus influencing directly the further development and applicability of Siemens control hardware. Another direct impact of the eFlows4HPC project will be the possibility to evaluate the reliability of the developed multiphysics models for real physical assets like large power transformers. These models have an inherent challenge of verification in the operation phase where, due to physical limitations, only few measurement/ sensor data are available. Therefore, a large amount of calculations has to be performed to get a sufficient level of accuracy, which is currently not possible. eFlows4HPC thus contributes significantly to the accuracy of the control models and thus to the overall robustness of the power transformers.

16. NGI: Realistic high-resolution numerical simulations of landslides, tsunamis, rockslides, and avalanches are fundamental to research and consulting activities at NGI. Assessing the uncertainty associated with numerical simulations for each of these natural hazards in both operational forecasting and long-term hazard analysis

requires the executions of a large number of calculations, searching exhaustive parameter spaces of relevant physical attributes. HPCWaaS will facilitate the advancement of NGI codes not currently dimensioned for HPC implementation, and provide a framework that simplifies the design of operational simulations for a wider number of hazard types not currently addressed in eFlows4HPC. NGI will exploit new HPCWaaS mechanisms developed in eFlows4HPC to make more systematic and efficient use of HPC in geotechnical and geohazard applications across the NGI project portfolio, to accommodate and exploit expected increases in both data volume and numerical model complexity in adjacent disciplines such as modelling landslide and snow avalanche dynamics.

2.2.3 Knowledge Management and Protection

The following considerations regarding Knowledge Management and Protection will apply in the project:
Protection of Results. The partners are aware that, once the project is complete, the results will be of great interest and benefit to academia and industry, not only on a European but also on a global level. Due consideration will be given to this issue, as well as to the needs of dissemination and communication, whereby confidential information will be protected while, at the same time, non-confidential information will be published in order to fulfil the project's dissemination goals. In the case of the code, the objective is to publish all produced code in the project as **open source.**
Ownership of Results. Bearing in mind the significance and importance of Intellectual Property Rights (IPR) in fostering innovation, the work plan for the dissemination and exploitation of the results foresees a continuous review of the important and relevant IP issues over the whole project duration. The management of IPR issues derived from the results (developed technologies, tools, manuals, dissemination materials and web page) will be managed by BSC, which as Technical Coordinator will serve also as IPR Manager. The partners will define all actions to be undertaken in the future to guarantee proper handling of IPRs. IPR will be managed according to the provisions in the Project Grant Agreement with the EC, as well as those in the Consortium Agreement (CA). The partners have agreed to use the DESCA 2020 model for the CA. During finalisation of the CA, foreground and background knowledge will be identified and access rules detailed, such as confidentiality, ownership of knowledge, patents and access rights. In general, all the knowledge, data and results deriving from the projects carried out by the beneficiaries will remain their exclusive property. With respect to project partners, results (from the project) shall be considered free of charge for partners that form the eFlows4HPC consortium within the scope and duration of the project. However, in case any of the results belong to one or more partners, those results shall be licensed to the rest of the partners under favourable conditions, to the extent necessary to enable these partners to exploit their own results.

Section 3. Implementation

3.1 Overall strategy of the workplan

eFlows4HPC will run for 36 months and comprises the following workpackages: **eFlows4HPC platform WPs** (WP1, WP2, WP3), **the Pillars' WPs** (WP4, WP5, WP6), and **transversal WPs** (WP7, WP8, WP9).

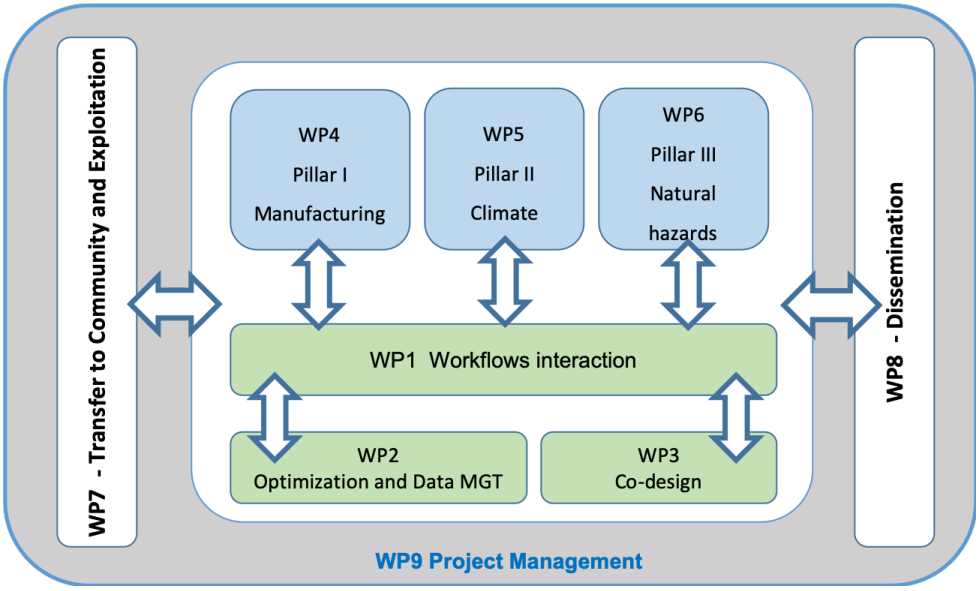


Figure 8: Graphical representation of the eFlows4HPC WPs and its inter-relations

The **eFlows4HPC platform WPs** will focus on providing the necessary software and resources to implement and efficiently execute the project target workflows. **WP1** will collect requirements, design the architecture and interfaces of the eFlows4HPC software stack, and perform integration and development activities to provide interfaces that support workflows that integrate HPC, HPDA and AI, as well as the design and implementation of the HPCWaaS methodology. **WP1 is centric** to the whole technological aspects of the project, and coordinates all technical activities with WP2 and WP3, and also with the Pillars' WPs (WP4, WP5, and WP6). These joint activities will be performed in a highly multidisciplinary team following a co-design approach between all WPs. With such goal, it also includes support for development activities in the Pillars. **WP2** will perform the necessary optimizations of the existing data analytics and ML libraries, as well as runtime systems to efficiently deploy, execute and manage resources and data. WP2 also has a special emphasis on data management, with specific tasks for storage technologies. It will implement a data logistics service, and deliver a data catalogue. **WP3** will identify specific kernels from the Pillars and optimize them for specific aspects related to the specialization for given architectures (the EPI, accelerators, heterogeneity, storage) and will also look into other optimization aspects of workflows (for example, towards reducing the energy required to execute them). **WP2 and WP3 tasks will contribute to the eFlows4HPC platform.**
The **Pillars' WPs** will focus on a set of target sectors (manufacturing, climate, natural hazards) and will demonstrate through real use cases how the technologies delivered by the generic software stack WPs can be used in dynamic and complex application workflows that integrate HPC simulation and modelling, data analytics and machine learning. **WP4** will focus on the generation of **reduced order models** for **digital twins** in manufacturing. **WP5** will focus on **dynamic and adaptive workflows for climate modelling**. **WP6** will focus on defining **workflows and policies for urgent computing for natural hazards**.
The results of the Pillars' WPs will not only be the specific optimized workflows for their target area topics but also **methodologies for the development of these types of workflows** for other similar problems. In this sense, it is very important to **transfer** these methodologies to the different **target communities**. This will be done in **WP7**, through the collaboration with the existing **HPC CoEs** and also through specific workshops organized to transfer the results. Finally, the **transversal WPs** will be responsible of transfer to communities and exploitation (WP7), support all activities related with dissemination and communication (**WP8**), and project and coordination management (**WP9**).

3.1.1 eFlows4HPC Work package List

WP No	WP Title	Lead Part. No & Short Name		PM	Start month	End month	Objectives addressed
1	Workflow interfaces for the integration of HPC, data analytics and ML	1	BSC	180	1	36	STO1, STO2 STO3, PSO1 SIO3
2	Optimization of runtime and libraries, workflow deployment, resource management, and data management	3	FZJ	143	3	36	STO1, STO3 STO4, SIO3
3	Co-design aspects between applications, software stack and actual hardware	4	UPV	108	3	36	STO1, STO5 SIO3
4	Pillar I: ROM: Digital twin in Manufacturing	2	CIMNE	193	1	36	PSO1, PSO2 SIO1
5	Pillar II: Dynamic and adaptive workflows for climate modelling	7	CMCC	102	1	36	PSO1, PSO3 PSO6
6	Pillar III: Urgent computing for natural hazards	1	BSC	149	1	36	PSO1, PSO4 PSO5, SIO2
7	Outreach to Communities and exploitation	5	ATOS	45	1	36	SIO1, SIO3 SIO4
8	Dissemination and Communication	1	BSC	26	1	36	SIO3, SIO4
9	Project Management	1	BSC	25	1	36	all
			Total PMs	**971**			

3.1.2 Detailed workplan description

The overall approach will be organized in **two iterations**, each of them **of two phases**, well mapped in terms of milestones and deliverables to the typical timing of the reviews of H2020 projects; Iteration 1 (Phases 1-2, M18) and Iteration 2 (Phases 3-4, M36).

Iteration 1:

Phase 1 (M1-M4): Requirements' analysis and architecture design: In this phase, the requirements from the Pillars will be collected and the architecture of the workflow stack will be defined. This phase will also have as objective the definition of metrics to evaluate and validate the workflows. Some initial internal training activities are planned during this phase.

Phase 2 (M5 - M18): First development iteration: In this phase the eFlows4HPC platform WPs will perform the corresponding integration and development activities to deliver at M12 the **first version of the eFlows4HPC platform**. The Pillars during this phase will perform integration and development activities to deliver at month 18 a **first version of the Pillar workflows**, including the evaluation of the metrics defined in Phase 1. The Pillars will also give **feedback** to WP1-WP3 at the end of this phase in order to update their requirements for phase 3. The activities between the different WPs are coordinated through WP1, which is the nexus link between WP1-WP6. The project plans to perform a hackathon across all WPs and Pillars by month 15, as a reinforcing training activity before the end of Iteration 1. This activity will include some external participants to disseminate the project activities to selected stakeholders. During this phase, the exploitation activities will perform a market analysis and release it together with the initial exploitation plan by M15.

Iteration 2:

Phase 3 (M19 - M30): Second development iteration: This phase is similar to the previous one, with the eFlows4HPC platform WPs delivering the **second version of the eFlows4HPC platform** at month 24 and the Pillars delivering the **second version of the workflows** at month 30. A **review of the requirements** is performed by WP1 (with the support of the Pillars) between phases 2 and 3, during months 16-19, to update them according to the results of the Pillars after the first development iteration. In this phase, the existing workflows, models and different project artifacts will populate the **workflows registry, model repository and libraries' catalogue** of the project, performing the initial tests of the **HPCWaaS methodology**. At the end of this phase, **external training and joint workshops** with the relevant CoEs (especially EXCELLERAT, ESiWACE and ChEESE) will be performed, in collaboration with the Focus CoE. In this second phase, the exploitation activities will define the exploitation strategy (M24).

Phase 4 (M31 - M36): Validation and final versions delivery: In this phase, partners will work on finalizing the software releases for the end of the project. Given the feedback provided by Pillars with the second version of the workflows, the eFlows4HPC platform WPs will work on the final details to improve and fix the **final release** by M33. The Pillars will perform a validation activity during this phase, including the evaluation of the metrics, releasing the **final version of the workflows** at M36. The last version of the different workflows, models and other project components will be released in the **workflows' registry, model repository and libraries catalogue**. A final exploitation report will be released at month 36.

The general dissemination and communication activities (WP8) and project management activities (WP9) will run during the whole project lifetime.

WP number	1										**Duration**			M01 – M36			
WP Title	Workflow interfaces for the integration of HPC, data analytics, and ML																
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	**BSC**	CIMNE	FZJ	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	**45**	0	30	18	47	28	12	0	0	0	0	0	0	0	0	0	**180**

Objectives

WP1 will comprise integration and development activities to provide interfaces to support workflows that integrate HPC, high performance data analytics and machine learning. More specifically, the WP objectives are:

- Definition of the requirements of the eFlows4HPC software stack. These requirements will be derived by the Pillars' use case needs.

- Definition of the eFlows4HPC architecture. The eFlows4HPC software stack will be composed of a set of existing components for workflow definition and orchestration, HPC, data analytics, and machine learning components, organized in different layers.
- Definition and implementation of the necessary interfaces between the different layers and components for its smooth integration.
- Definition, implementation, and deployment of the project workflows registry and models repository.

Additionally to these goals, WP1 is centric in the project idea, and it is the glue that coordinates and integrates all technical work. In this sense, it will integrate results from WP2 and WP3 that will feed the Pillars' developments. Also, it will perform co-design activities, together with the Pillars', following a multidisciplinary approach, both in the definition of requirements and metrics, and support for the definition and development of the devised Pillars' solutions.

Description of work

The WP is organized in six tasks. Task 1.1 will define, together with tasks 4.1, 5.1 and 6.1, the Pillars' requirements towards the eFlows4HPC platform and infrastructure, as well as the metrics that will be used on the evaluation of the Pillars' results. Task 1.2 will define the architecture of the eFlows4HPC software stack and task 1.3 will evolve this architecture by defining all the necessary interfaces between the different components. Task 1.4 will focus on the development of the corresponding interfaces defined in task 1.3, and also integrates the results of tasks in WP2 to deliver the eFlows4HPC software stack. Task 1.5 will design and develop the HPCWaaS methodology, central element of the project platform, with the goal of widening the access of HPC to newcomers. Task 1.6 will perform activities to create the different repositories and catalogues delivered by the project. Task 1.7 aims at supporting the Pillars on their developments that require the use of data analytics and machine learning libraries, and workflow environments, following a co-design approach. The tasks in this WP will deliver two prototypes, one at month 12 and another at month 24, to enable the two-phase scheme proposed for the Pillars.

Task 1.1 Requirements analysis from Pillars on the eFlows4HPC software stack and on the heterogeneous infrastructures system (M1 – M4, M16-M1) (BSC, all)

This task will collect the different requirements from the Pillars' use cases for the eFlows4HPC software stack and for the heterogeneous infrastructure system. This task is done in collaboration with tasks 4.1, 5.1 and 6.1 with the purpose of deriving the requirements that are foreseen from the use cases on the eFlows4HPC software stack, on the HPCWaaS deployment and execution methodologies and on the underlying heterogeneous infrastructure. Between the type of requirements that will be collected, the task will consider the file and data formats used by the different use cases to unify how data is exchanged in the different steps of the workflows. The requirements will be collected in two phases: a first phase at the beginning of the project, with the initial requirements of the Pillars, and a second phase, half-way of the project, when the Pillars should be able to give more detailed requirements for the second phase of the eFlows4HPC software stack. The first phase of this task will also define metrics on accessibility, scalability, performance, energy efficiency, reliability, and cost.

Task 1.2 Architecture design of the eFlows4HPC software stack (M1-M4, M16-M19) (BSC, Atos, FZJ, DtoK Lab, CMCC, UPV)

This task will design the architecture of the eFlows4HPC software stack. We have devised an initial architecture, which is described in section 1.3.1. The envisaged software stack is composed of the integration of existing software components which are organized in different layers. The task, based on this initial architecture and taking into account the discussions in task 1.1 will derive the actual architecture. Since the requirements will be revised after the first half of the project, the architecture may also be subject of some variations and a review is also foreseen before the second phase of the project.

Task 1.3 Definition of the interfaces for the integration of HPC, HPDA and AI components (M5-M7, M20-M21) (DtoK, BSC, Atos, FZJ, CMCC, UPV)

This task will focus on the design of the necessary interfaces for the integration of HPC, data analytics and machine learning in a single workflow. Given the architecture defined in task 1.2, this task will carefully take into account the necessary data-driven and control-driven interactions between the different layers and components and will define the interfaces required for a smooth execution of the workflows. Similarly to tasks 1.1 and 1.2, this task will be performed in two phases.

Task 1.4 Software stack integration (M8-M33) (BSC, Atos, FZJ, CMCC, UPV, DtoK)

This task will implement the corresponding developments in the software stack components in order to implement the interfaces designed in task 1.3. The correct integration of the different components will be validated with a set of integration tests following a continuous development and integration methodology. Results of tasks in WP2 are also integrated here, and will also be subject of the test and integration development methodology. The different releases of the software stack performed by this task will also include deployments in the target project infrastructures.

Task 1.5 Widening access to HPC systems: HPC Workflow as a Service (HPCWaaS) (M8– M24) (Atos, BSC)
The task will focus on widening the access to HPC to newcomers, and to facilitate the deployment and execution of complex workflows in HPC systems. To this end, the task will design and implement a mechanism to deploy HPC Workflows as a Service (HPCWaaS). The mechanism should enable workflow developers to register workflows they designed. These registered workflows can be later deployed and executed by new users. The initial design will leverage from Ystia orchestrator (Atos) based on the TOSCA standard. The task will need to design and implement the workflow registry and the integration of the Ystia orchestrator into the eFlows4HPC software stack to deploy and execute the workflows composed of HPC, data analytics and ML components. This may imply extending the Ystia orchestrator with features like dynamic workflows, failover workflows, data-aware workflows. Additionally to workflow instances, workflow developers will also be able to register workflow templates, which are generic workflow descriptions that should be customized to match the requirements of a specific use case. In this sense, the environment will also allow to tailor the workflow templates to specific instances and to deploy them for execution.

Task 1.6 Creation of artifacts public repositories (M20 – M33) (FZJ, BSC, Atos, CMCC, UPV, DtoK)
The task will create a workflow registry for workflow instances and templates, a model repository, and a libraries' catalogue, that could be offered to the different user communities. It will focus on the actual implementation of these registries, repositories, and catalogue and on its population with project results. The model repository will be complemented with AI models management features covering typical ML/DL models lifecycle like training, tuning, inference capabilities. The task will support the Pillars' on the generation of the corresponding images and artifacts to populate them.

Task 1.7 Development support for Pillars' use cases (M4-M36) (UPV, BSC, Atos, FZJ, CMCC, DtoK)
This task will focus on the support required to implement the different Pillars' use cases, including the use of data analytics, machine learning libraries, and workflow environments. While most of the developments of the Pillars' use cases will be performed in WP4, WP5 and WP6, those developments or support that correspond to the eFlows4HPC software task will be performed in WP1, WP2 and WP3. The activities performed in this task will involve the design and initial implementation of the solutions, while task 2.1 will focus on the runtime aspects and optimization of the solutions for the Pillars. Example of the activities performed in this task include the support on the design and development of the solutions, day-to-day support on questions, bugs, issues with the different libraries and tools, etc.

Deliverables

D1.1 Requirements, metrics and architecture design (BSC, R, PU) [M4]. First version of the pillar's requirements, evaluation metrics and architecture design (output from tasks 1.1, 1.2, 4.1, 5.1, 6.1)

D1.2 eFlows4HPC interfaces and Iteration 1 software stack release (DtoK, R+OTHER, PU) [**M12**]. First version of the design and implementation of the eFlows4HPC software stack interfaces. Tasks 1.3, 1.4, 1.5 contribute to this deliverable.

D1.3 Revision of Requirements and architecture design (BSC, R, PU) [M18]. Second version of the Pillars' requirements and architecture design (output from tasks 1.1, 1.2, 4.1, 5.1, 6.1)

D1.4 eFlows4HPC interfaces and Iteration 2 software stack release (Atos, R+OTHER, PU) [**M24**]. Second version of the design and implementation of the eFlows4HPC software stack interfaces. Tasks 1.3, 1.4, 1.5, and 1.6 contribute to this deliverable.

D1.5 eFlows4HPC interfaces and final software stack release (BSC, R+OTHER, PU) [**M33].** Second version of the design and implementation of the eFlows4HPC software stack interfaces. Tasks 1.3, 1.4, 1.5, 1.6 and 1.7 contribute to this deliverable.

WP number	2								**Duration**					M03 – M36			
WP Title	Optimization of runtime and libraries, workflow deployment, resource management, and data management																
Participant number	1	2	**3**	4	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	BSC	CIMNE	**FZJ**	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	36	0	**47**	8	20	20	12	0	0	0	0	0	0	0	0	0	**143**

Objectives

- Optimize the eFlows4HPC machine learning and workflow runtimes to obtain the required levels of performance and energy efficiency.
- Enable a stable deployment platform, supporting the selected container solution for dynamic deployment of the HPC Workflow-as-a-Service concept.
- Provide a stable set of innovative storage solutions to support scientific workflows.
- Identify relevant data sources, list them in a Data Catalogue and into an Integrated Data Logistics Service.
- Implement Data Pipelines to fuel Pillars use cases.

The implementations performed in the tasks of this WP will be integrated with the results of task 1.4 and into its deliverables.

Description of work

The goal of this work package is to conduct, supervise, and optimize the efficient and flexible deployment of the integrated software stack defined in WP1, as well as the required support for the implementation of the use cases from the Pillar's WPs in a co-design manner. The three main activities areas of this work package are the optimization of the existing components in the eFlows4HPC platform (machine learning and data analytic libraries, workflow runtime, etc.), the efficient deployment of services and parts of the workflows (i.e., computation support) and provision of data resources in terms of both storage (T2.3) and relevant data items (T2.4). The tasks in this WP will deliver two prototypes, one at month 12 and another 24 to enable the two-phases scheme proposed for the Pillars.

Task 2.1 Data analytics, machine learning libraries and workflow system optimizations (M3 – M33) (DtoK, BSC, UPV, CMCC, Atos)

This task will develop the necessary optimizations to obtain a performant, scalable and energy efficient eFlows4HPC platform. With this objective, the different components of the eFlows4HPC software stack will be enhanced in this task towards the project goals and taking into account the Pillars' requirements derived in tasks 1.1, 4.1, 5.1, and 6. More specifically, this task will involve optimizations in the data analytics and machine learning libraries, in the workflow runtimes and orchestrator engines.

Task 2.2 Deployment optimization through container methodology (M3 – M24) (Atos, FZJ, BSC)

To simplify, speed-up and improve the overall flexibility of the deployment, the project will provide the different components of both the software stack and the workflows in the form of container images. To this end, this task will evaluate available solutions, and subsequently derive a methodology to encapsulate and deliver the different project components with containers. This task will work in collaboration with task 1.5 toward the deployment of the HPCWaaS methodology.

Task 2.3 Storage technologies (M3 – M36) (BSC, CMCC)

This task will optimize the storage related aspects of the software stack. In this sense, this task will focus on alternative solutions to the traditional file systems, such as persistent and shared storage libraries (dataClay, Hecuba, Ophidia and Qbeast). The task will focus on the optimization and extensions required on these technologies to fulfill the pillar use cases requirements.

Task 2.4 Data logistics service and transfer strategies and its integration with the eFlows4HPC stack (M3 – M36) (FZJ, BSC)

Regardless of the approach used by the scientists (HPC, High Performance Data Analytics), they often require lots of high-quality data, timely delivered to the processing facilities to fuel the planned scientific endeavours. This task will develop and deploy a Data Logistic Service to help to move and pre-process data. The idea is to advance the current manual approach for data movement towards on-demand, self-service, automatic approach. The initial prototype of the service will be based on Juelich Data Logistic service which employs Apache Airflow as underlying technology. Depending on the feedback and requirements provided in the co-design process with the Pillars, the service can be extended to enable real-time streaming-based solution based on technologies like Apache Kafka. The service will have to be integrated with source services offering the required data on one hand and with storage technologies from T2.3 on the other. The first outcome of this task will be a Data Catalogue (D2.1) that will summarize all data sources used by the Pillars, including their API schemas, protocols used, and intended usage. The Data Catalogue will be used to prioritize the integration effort of the Logistics Service. The task will be actively providing parts of the eFlows4HPC platform in MS2, MS4, and MS6.

Deliverables

D2.1 Data Catalogue (FZJ, R+OTHER, PU) [M6]. Based on the requirements report D1.1 this deliverable will analyse and describe the data sources used by the Pillars. This information will be made available in the form of an electronic document or service.

D2.2 Technology Evaluation, Containerization and Optimization Strategy (Atos, R, PU)[M8]. Based on the available requirements, this deliverable will derive a strategy for optimizing deployment in all envisioned

dimensions: by optimizing libraries and runtimes, using containers, and application of emerging storage solutions. Tasks 2.1, 2.2, 2.3 contribute to this deliverable.
D2.3 First version of Data Logistics (FZJ, R+OTHER, PU)[M14]. This deliverable will be the first production version of Data Logistics service integration with selected storage technology and demonstration of a data pipeline motivated by the Pillars' use cases. Task 2.3 and 2.4 contribute to this deliverable.
D2.4 Report on Implementing Containerization and Optimization Strategy (Atos, R+OTHER, PU)[M28]. This deliverable will describe the implementation of the Optimization Strategy (D2.2) and possibly reevaluate the approach taken. Tasks 2.1, 2.2, 2.3 contribute to this deliverable.
D2.5 Final Report on Data Logistics Implementation (FZJ, R+OTHER, PU)[M36]. This report will comprise the description of the final version of Data Logistics Service, implementation of data pipelines motivated by Pillars. Task 2.3 and 2.4 contribute to this deliverable.
Additionally to these deliverables, tasks in WP2 contribute to the software releases included in deliverables D1.2, D1.3 and D1.4.

WP number	3									**Duration**				M03 – M36			
WP Title	Co-design aspects between applications, software stack and actual hardware																
Participant number	1	2	3	**4**	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	BSC	CIMNE	FZJ	**UPV**	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	24	0	9	**40**	0	22	13	0	0	0	0	0	0	0	0	0	**108**

Objectives
WP3 will set as objectives:

- The identification of specific parts (kernels) of the application Pillars under consideration, in order to improve key performance indicators such as energy consumption and execution time.
- Assessing the optimization of such kernels and their customization to specific heterogeneous architectures, including GPUs, FPGAs and custom accelerators, with special emphasis on the EPI.
- The investigation of specific kernels to improve energy efficiency by trading off accuracy/precision and energy/performance when possible.
- The investigation of how the storage components of the software stack will benefit from novel storage technologies to improve the performance of the Pillars of the project.
- The specification of metrics that will enable to evaluate the achieved improvements with the project results.

The optimized kernels, derived from the WP, will be prepared for integration within the Pillars' workflows.

Description of work
WP3 will work at two levels. First, it will operate between the applications and the eFlows4HPC software stack, identifying optimization opportunities (task 3.1). In addition, it will act between eFlows4HPC software stack and hardware by designing new implementation opportunities derived from the use of heterogeneous architectures (tasks 3.2, 3.3, 3.4) and new storage approaches (task 3.5). Evaluation will be performed in task 3.6. Some tasks in this WP will deliver inputs for the second prototype of the eFlows4HPC at M24 and some optimized kernels to the Pillars' workflows at M33.
Task 3.1 Identification of application and software stack bottlenecks that can be optimized with new storage, computing and communication technologies (M3 – M9) (UPV, DtoK, BSC, CMCC)
This task will investigate and identify application bottlenecks in the Pillars, as well as in the software stack components, that prevent further scalability in terms of performance and energy efficiency. current and new heterogeneous computing components will be considered in this task, performing initial analysis of optimization opportunities. The impact of the communication network will be considered and potential optimizations will be identified. For storage (and its associated impact on communication), issues such as data-mapping awareness, data locality, task synchronization, and tasks-to-processors mapping will be analysed in the pillar applications and approaches for improving performance and speed-up of the designed software stack will be analysed. The task will be fed with the outcome of tasks 1.1, 4.1, 5.1, 6.1 and 1.2.

Task 3.2 Adaptation of identified application kernels to heterogeneous components (GPU, FPGA, custom accelerators) (M9 – M33) (UPV, BSC, CMCC)
This task will deal with heterogeneous components such as FPGAs, GPUs and custom accelerators. The task will target the kernels identified in T3.1 and will optimize their implementation for performance and energy efficiency goals. Specific optimizations that adapt (reduce) precision in the intermediate computations while maintaining the accuracy of the result will be considered as well. At M24, an initial set of kernels will be provided to the eFlows4HPC second prototype. At M33, optimized kernels will be provided to the Pillars' workflows.
Task 3.3 Adaptation of identified application kernels to EPI (M9 – M36) (BSC, UPV)
This task will select a subset of application kernels from those identified in T3.1 and will evaluate their performance in the EPI. This evaluation can be done with two alternative platforms: with the MareNostrum Exascale Emulation Platform (MEEP), which will develop an FPGA-based system that will emulate the EPI accelerator; or in a real platform in the Fujitsu post-K cluster that will be installed at BSC in 2020. This cluster will be based on the ARM processor, which has a very similar architecture to the EPI. The MANGO prototype will be also targeted for emulation of the EPI. At M24, an initial set of kernels will be provided to the eFlows4HPC second prototype. At M33, optimized kernels will be provided to the Pillars' workflows.
Task 3.4 Optimization of Machine Learning kernels in heterogeneous components (GPU, FPGA, custom accelerators) (M9 - M33) (UPV, BSC, CMCC)
In this task, machine learning algorithms implemented in the project will be assessed and optimized for heterogeneous components. The algorithms will be analysed and optimized for energy efficiency and performance. At M24, the kernels will be provided to the eFlows4HPC second prototype. At M33, the kernels will be provided to the Pillars' workflows.
Task 3.5 Optimization of data management with new storage technologies (M9 – M33) (BSC, CMCC)
This task will evaluate new storage technologies, such as Intel OptaneDC persistent memories, so that the components of the eFlows4HPC software stack can benefit from a storage hierarchy composed of a mix of traditional storage devices and novel storage devices. In particular, persistent memories can be used in different modes: playing the role of a disk (although faster and smaller) or acting as RAM (but bigger and slower). This task will work on the optimization of the storage bottlenecks identified in T3.1, using Hecuba and dataClay as well as on Ophidia to transparently manage data on persistent memories in the most appropriate way.
Task 3.6 Evaluation of applications towards achieving effective improvement on measurable metrics (M9 – M33) (UPV, BSC, DtoK Lab)
This task will aggregate the evaluation of the workflow when using the identified metrics from task 1.1, and by adopting the optimized kernels and storage technologies when using the targeted heterogeneous architectures. An initial evaluation will be performed at M12. In addition, at M18 and M33, the task will provide evaluation outputs for the eFlows4HPC prototype.

Deliverables
D3.1 Application bottlenecks and optimization opportunities on heterogeneous components (UPV, R, PU) [M9]. This report will include the identified bottlenecks as well as a preliminary analysis of such issues when considering the communication architecture, the storage architecture, and the use of heterogeneous components. This report will be the baseline for the work of the remaining tasks in this WP. This deliverable will collect the outcome from task T3.1.
D3.2 Initial draft of optimized kernels for heterogeneous components (UPV, R, PU) [M18]. This document will reflect a first draft version of the optimized kernels developed in the project for the use of heterogeneous components. It will compound application kernels and neural network kernels. This deliverable will collect the work from tasks T3.2 and T3.4, including associated evaluation work from T3.6.
D3.3 Initial draft of optimized kernels for EPI (BSC, R, PU) [M18]. In this report the draft version of the optimized kernels for the EPI will be described. Their implementation on the emulated hardware platform will be provided and its projected performance and energy saving. This deliverable will collect the work from task T3.3, including associated evaluation work from T3.6.
D3.4 Optimized kernels for heterogeneous components (UPV, R, PU) [M33]. This document will reflect the final version of the optimized kernels developed in the project for the use of heterogeneous components. It will compound application kernels and neural network kernels. This deliverable will collect the work from tasks T3.2 and T3.4, including associated evaluation work from T3.6.
D3.5 Optimized data management with new storage technologies (BSC, R, PU) [M33]. This document will reflect the optimizations performed with new storage technologies. This deliverable will collect the work from task T3.5 and associated evaluation work from T3.6.
D3.6 Optimized kernels for EPI (BSC, R, PU) [M36]. In this report the optimized kernels for the EPI will be described. Their implementation on the emulated hardware platform will be provided and its projected

performance and energy saving. This deliverable will collect the work from task T3.3, including associated evaluation work from T3.6.

WP number	4									**Duration**				M01 – M36			
WP Title	Pillar I: ROM: Digital twin in Manufacturing																
Participant number	1	**2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	BSC	**CIMNE**	FZJ	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	0	**81**	0	0	0	0	0	43	43	0	0	0	0	0	26	0	**193**

Objectives

WP4 aims at:

- Developing and demonstrating state-of-the-art simulation techniques for the development of "Digital Twins" of manufactured objects.
- Integrate Model Order Reduction approaches with the eFlows4HPC workflow stack
- Enable both training and inference to take advantage of effective HPC usage.
- Make use of automatic remeshing technology and immersed simulation techniques to allow automatizing the construction of complex simulations.

Description of work

Aside of the identification of the requirements, to be performed in collaboration with other WPs (WP1), WP4 is organized around a number of use cases to be identified in collaboration with the industrial partner (task 4.2). Such use cases will be used to develop and deploy the software infrastructure needed to reach the final goal (task4.3) and later to develop and validate the solution for the use cases (tasks 4.4 and 4.5). The WP is concluded by the integration of the technology in the HPCWaaS workflow repository (task 4.6).

Task 4.1 Requirements from Pillar I on the eFlows4HPC software stack and heterogeneous infrastructure system (M1 – M4, M16-M19) (CIMNE, Siemenes, SISSA, INRIA)

This task will develop detailed requirements from the hardware and software points of view in order to successfully integrate the solvers with the eFLows4HPC workflow stack. The task will also define data exchange mechanisms so as to ease future collaboration between the different teams as well as define metrics on accessibility, scalability, performance, energy efficiency, reliability, and cost. This task will coordinate with task 1.1.

Task 4.2 Design of the Pillar I use cases (M1 – M6) (Siemens, CIMNE, SISSA, INRIA) The partners will jointly identify relevant use cases. The main goal will be to demonstrate the potential of simulation-based Digital Twins in the real time simulation of manufactured products with the goal of exploring future integration in industrial design workflows. This task will target both the definition of relevant applications to be targeted for ROM construction and the identification of use cases for ML models to be built on the top of the newly developed models

Task 4.3 Iteration 1 of development of the Pillar I use cases (M7 – M18) (SISSA, CIMNE, INRIA)

The task will put in place the tools needed for data mining of Full Order Models. Both linear (SVD based) techniques and nonlinear ROM techniques will be considered in this phase. The tools will be made available as prototypes and then considered for inclusion within the dislib library

Task 4.4 Iteration 2 of development of the Pillar I use cases (M19 – M30) (CIMNE, Siemens, INRIA, SISSA)

Based on the developments of Iteration 1, this task will focus on the development of realistic reduced order models for the selected use cases. The models will also be prepared for interaction with machine learning, and to perform robustness tests and or for automatic model calibration. This is particularly crucial since sensor data input is known to be spotty and thus requires simulation to fill the gaps. Mesh adaptivity will be introduced in the workflow as a part of the task.

Task 4.5 Validation of the Pillar I use cases (M30 – M36) (Siemens, CIMNE, SISSA, INRIA)

The development ROM models will be verified in lab, ideally by comparison to sensor data. The task will also evaluate the developed use cases with the metrics defined in task 4.1.

Task 4.6 Integration of results in workflow repository and HPCWaaS interface (M30 – M36) (BSC, CIMNE, INRIA, SISSA)

Integration will be achieved by following WP1 guidelines (task 1.6), with the goal of maximizing usability of the developed tools. The task will also explore models to ensure the financial sustainability of the project beyond the time scale of the project

Deliverables

D4.1 Requirements on the eFlows4HPC software stack from Pillar I and evaluation metrics (CIMNE, R, PU) [M4]. Summary of conclusions on the pillar I requirements.

D4.2 Design of the Pillar I use cases (Siemens, R, PU) [M6]. Report on the choice of industrially relevant problems to be addressed in the WP. Report will identify success criteria as well as potential breakthroughs.

D4.3 ROM Tools Release (SISSA, OTHER, PU) [M18]. Code Release of essential tools for ROM preparation.

D4.4 Demo ROM (CIMNE, OTHER, PU) [M30]. Release of demonstrator ROM model.

D4.5 Release of HPCWaaS integrated solver stack (BSC, OTHER, PU) [M33]. Release of software stack integrated in the HPCWaaS interface

D4.6 Validation of the Pillar I use cases (Siemens, R, PU) [M36]. Report on the validation of the ROM models. Evaluation of success criteria.

WP number	5			**Duration**										M01 – M36			
WP Title	Pillar II: Dynamic and adaptive workflows for climate modelling																
Participant number	1	2	3	4	5	6	**7**	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	BSC	CIMNE	FZJ	UPV	ATOS	DtoK	**CMCC**	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	36	0	0	0	0	0	32	0	0	0	0	0	34	0	0	0	**102**

Objectives

This Work Package (Pillar II) aims to demonstrate the applicability and effectiveness of the eFlows4HPC approach and technologies with a set of key use cases in the climate modelling domain.

More specifically WP5 sets as objectives:

- Development of intelligent and novel end-to-end ensemble Earth System Modelling (ESM) workflows able to (i) rapidly adapt and evolve according to the dynamic conditions of the climate simulations, and (ii) make a better use of computational and storage resources by performing a smart (AI-driven) pruning of ensemble members (and releasing resources accordingly) at runtime;
- Seamless, transparent and efficient integration of different components of the ESM workflow (from simulation, to post-processing, HPDA and learning) into the same experiment, overcoming current gaps and barriers;
- Development of data-driven ML/DL models to help understanding and grasping key features as well as to produce added-value products (i.e. Tropical Cyclones track) from the climate simulations;
- Investigation of post-processing versus in-situ approaches to study how they respond to the scientific needs, and to the use of computational and storage resources;
- Evaluation of data-intensive versus data-driven approaches with respect to key features (like TC track) by performing a scientific multidimensional analysis of (i) accuracy, (ii) uncertainty, (iii) time-to-solution, and (iv) energy consumption.

An agile methodology will be adopted to ensure the right level of flexibility during the entire software development process.

Description of work

The overarching goal of WP5 is to take advantage of the eFlows4HPC architecture to enhance innovation for intelligent and integrated end-to-end HPDA-enabled ensemble ESM workflows. The initial requirements analysis will be performed in Task 5.1, while the workflow design will take place in Task 5.2. Two implementation phases (Task 5.3 and Task 5.4) during the project will address the development of two different software releases for the selected use cases, which will also integrate core components from WP1, WP2 and WP3. Task 5.5 will provide a comprehensive scientific validation of the proposed solution, while Task 5.6 will ensure that the final workflows/models of the use cases will be integrated into the eFlows4HPC repository and with the HPCWaaS

interface. An agile approach will be adopted to ensure the right level of flexibility during the entire software development process.

Task 5.1 Requirements analysis from Pillar II on the eFlows4HPC software stack and heterogeneous infrastructure system (M1 – M4) (AWI, CMCC, BSC)

This task represents the first step in the software development process of the proposed climate modelling pillar workflows. It will perform a formal identification, gathering, analysis and prioritization of requirements for application design and implementation. Pillar II will target innovative, dynamic and adaptive ensemble ESM workflows for climate simulations, integrating HPDA and ML/DL components to efficiently extract added-value features. State of the art data-intensive versus novel data-driven approaches will be considered; and both post-simulation and in-situ analytics use cases will be investigated. The requirements analysis process will lead to a report (D5.1) with a comprehensive list of requirements, properly documented use cases and scenarios. It will also contribute to the eFlows4HPC Milestone1. The task will also define metrics quantifying accessibility, scalability, performance, energy efficiency, reliability, and cost that will be incorporated in the report (D5.1). This task will coordinate with Task 1.1.

Task 5.2 Design of the Pillar II use cases (M1 – M6) (BSC, CMCC, AWI)

This task will deliver the design of the selected Pillar II use cases. It will take into account the requirements coming from Task 5.1 and deliver thorough documentation (D5.2) in terms of architecture, modules, standards, data structures, workflows and APIs.

Task 5.3 Iteration 1 of development of the Pillar II use cases (M7 – M18) (CMCC, BSC, AWI)

This task is related to the Phase 1 implementation of the Pillar II use cases. It will deliver a first release of the modules needed by the use cases, including a first level of integration of the components.

In particular, during this phase the following activities will be carried out:

- Initial development of the workflow to enable dynamic and adaptive ensemble simulations;
- Preliminary integration of the workflow support into small-scale production-level settings. To this end, a selected set of ESMs (i.e. OpenIFS-FESOM2, CMCC-CM2) will be considered;
- First implementation of the ML/DL models to extract added-value climate products, like Tropical Cyclone (TC) tracks, and associated evaluation on real simulation outputs.
- Preliminary insights from the multidimensional analysis comparing data-intensive approaches versus data-driven ones.

An initial release of the different software components for the use cases will be made available as open source; and some prototypes for evaluation and feedback from the end-users will be set up too.

A first integration activity with the core components of eFlows4HPC provided by WP1, WP2 and WP3 will be performed and technical feedback will be reported to the developers across the whole software stack. The first release of the software components will be properly documented as part of the activities of this task into D5.3 and will contribute to the general eFlows4HPC Milestone3.

Task 5.4 Iteration 2 of development of the Pillar II use cases (M19 – M30) (CMCC, BSC, AWI)

This task is related to the implementation of the Pillar II use cases (Iteration 2). It will deliver the second release of the climate use cases with a stronger integration effort, thus delivering results across the entire ESM workflows.

In particular, during this phase the following activities will be addressed:

- Second release of the AI-assisted workflows support (e.g. through ML models) for ensemble ESM simulations;
- Second release of the DL/ML models to extract added-value climate products (e.g. TC track);
- Integration of the different modules into production-level end-to-end ESM workflows settings and multidimensional analysis of the proposed solution.

This task will also explore how post-processing versus in-situ approaches respond to the scientific needs and to the use of computational and storage resources.

The second release of the different components will contribute to the eFlows4HPC Milestone4 and it will be properly documented into D5.4. The software will be made available as open source for end-user (climate modelling groups, Task 5.5) evaluation and the integration activities planned in Task 5.6.

Task 5.5 Validation of the Pillar II use cases (M30 – M36) (AWI, CMCC, BSC)

This task will assess whether requirements identified in Task 5.1 have been fulfilled. The experience on the usage of the components will determine the suitability of the eFlows4HPC solution. The validation will include several assessments; it will be both internal and external; it will be based on different metrics defined in Task 5.1 of the fulfilment of requirements, completeness and appropriateness of the implementation and will include a comprehensive scientific evaluation of the results produced, as well on metrics related to accessibility, scalability, performance, energy efficiency, reliability and cost. Additionally, lessons learnt from the selected ESMs and

general guidelines about the applicability of the results to other ESMs available in the community, will be provided. A consolidated report (D5.5) will be prepared and delivered as output of this task.

Task 5.6 Integration of results in workflow repository and HPCWaaS interface (M30 – M36) (BSC, CMCC, AWI)

This task will integrate the use cases components, workflows and models developed in Task 5.4 with the eFlows4HPC workflow repository and the HPCWaaS interface. Such integration will require a strong interaction with the end-users to get validation feedback and shape the final version of the Pillar II use cases. The task results will be documented and delivered as part of the eFlows4HPC Milestone7.

Deliverables

D5.1 Requirements on the eFlows4HPC software stack from Pillar II and evaluation metrics. (AWI, R, PU) [M4] This report will include a prioritized list of requirements that will guide the design and implementation of the different Pillar II use cases as well as of the core eFlows4HPC components. The deliverable will also include a set of metrics to be used in the evaluation of the developed workflows.

D5.2 Design of the Pillar II use cases (BSC, R, PU) [M6] This report provides a complete design and comprehensive documentation of the software architecture of the Pillar II use cases.

D5.3 Pillar II - Iteration 1 Software Release (CMCC, OTHER, PU) [M18] This deliverable relates to the software and documentation released at the end of Phase 1 for the implementation of the Pillar II use cases.

D5.4 Pillar II - Iteration 2 Software Release (CMCC, OTHER, PU) [M30] This deliverable relates to the software and documentation released at the end of Iteration 2 for the implementation of the Pillar II use cases.

D5.5 Validation of requirements (AWI, R, PU) [M36] This report will include the process and outcome of the validation of the requirements (internal and external evaluations).

WP number	6			**Duration**							M01 – M36						
WP Title	Pillar III: Urgent computing for natural hazards																
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	**BSC**	CIMNE	FZJ	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	**36**	0	0	0	0	0	0	0	0	16	32	34	0	17	0	15	**149**

Objectives

WP6 aims to improve and optimize workflows related to urgent computing simulations for natural disasters, focusing on tsunamis and earthquakes. The main goal is to deliver faster end-to-end runs, with more robust and reliable workflows, and outcomes more usable to potential end-users. This implies a deep analysis of workflow dependencies, bottlenecks and resource management. A strong focus will be put toward integrating multiple sources of data and seamless interaction of the workflows into existing end-user decision flows. A comparison between the original workflows and those developed/improved within the project should be used to showcase the capabilities of the eFlows4HPC software stack.

Description of work

The main interest of this WP is making use of the eFlows4HPC components to boost urgent computing simulations for earthquakes and tsunamis. We will first analyse the requirements of the PTF and UCIS4EQ in task 6.1. The following effort will focus on building a sufficiently complete set of use-cases together with end-users in task 6.2. The developments associated to Iteration 1 and 2 have been split into two topical tasks: task 6.3 devoted to tsunamis and task 6.4 devoted to earthquakes. Parallel to the development, we will build recommendations regarding urgent access to HPC resources in task 6.7. A validation stage will follow the developments in Phase 3 in task 6.5 in order to assess the potential impact of the technologies developed. Simultaneously we will perform the final integration of the workflow in current operational environments (task 6.6).

Task 6.1 Requirements from Pillar III on the eFlows4HPC software stack and heterogeneous infrastructure system (M1 – M4) (BSC, UMA, INGV)

The task will perform a thorough analysis of present capabilities and future requirements of T-HySEA and the complete workflows for Faster Than Real Time (FTRT), Probabilistic Tsunami Forecasting (PTF) as well as for the UCIS4EQ workflow. Among the capabilities that would benefit from enhancements are resilience to failure

of the workflow environment, real-time workflow interaction, heterogeneous compute and edge compute capabilities. We could also benefit from precision/speed trade-offs from ROM or ML techniques when exploring the parameter space in uncertainty quantification. The required components will use data from task 6.2 and fuel tasks 6.3 and 6.4. The first phase of this task will also define metrics on accessibility, scalability, performance, energy efficiency, reliability, and cost. This task will perform its activities in coordination with task 1.1.

Task 6.2 Design of the Pillar III use cases (M1 – M6) (INGV, BSC, UMA, ETHZ, NGI, PSNC)

The task will gather a collection of historical seismic events of relevance due to their impact, challenging nature and representativity. All information regarding models (bathymetry, velocity and density models) and data will be collected, verified and made available for usage in tasks 6.3-6.5. The task may consider hypothetical events as well, to be used as drills to evaluate issues related to the operational capabilities of the urgent simulation workflows as well as to help designing policies in task 6.7. Ideally, there should be available state-of-the-art solutions to the aforementioned events, i.e. earthquake source data, shakemaps, and offshore tsunami wave observations for the events in urgent situations with current software packages.

Task 6.3 Development of the Pillar III workflows: earthquakes (M7 – M30) (ETHZ, BSC, INGV)

The focus of this task is twofold: (1) the development of novel features for UCIS4EQ as well as features that can be added as modules to it; and (2) generation of a precomputed database of local 3D velocity models. (1) will address the capacity of the workflow to be resilient to multiple types of failures, managing resources and tasks dynamically, building a real-time monitoring system to analyse the status of a run and making the results more usable to potential end-users, e.g. by adding ensemble simulation capabilities. (2) will involve full waveform inversions for a range of regions in order to create a collection of precomputed Earth models ready for urgent simulations.

At Iteration 1, Phase 2 we will port most components of UCIS4EQ to PyCOMPSs and use the core components of eFlows4HPC for a new release of the application. We will also produce new Earth models for the regions of interest defined in task 6.2. We will also develop an interface to support ensemble runs within UCIS4EQ, as a prototype. At Iteration 2, Phase 3 we will include the developments from WP1-3 to improve the resilience and efficiency of the code and enhance the workflow's monitoring capabilities. We will also enhance the uncertainty treatment, by including the previously built high-resolution models to enrich the ensemble for treating uncertainty in urgent computing.

Task 6.4 Development of the Pillar III workflows: tsunamis (M7 – M30) (UMA, INGV, NGI)

This task adapts an existing workflow for Probabilistic Tsunami Forecasting (PTF) consisting of the three stages: (i) seismic source ensemble initialization, (ii) multiple Faster Than Real-Time (FTRT) numerical tsunami simulations using the multi-GPU software T-HySEA, and (iii) post-processing with hazard aggregation enabling tsunami forecast. eFlows4HPC will allow a dynamic and iterative approach to the PTF workflow, allowing significant added functionalities: (1) a near-real-time source-estimation manager to initialize PTF ensemble forecast based on earthquake data, (2) an efficient cloud storage of model outputs to enable deeper exploitation of simulation results, and (3) increased event diagnostics. In addition, we will assess the possibility of updating PTF as other data are made available (e.g. moment tensor solutions, tsunami wave gauge data etc.). Subtask (1) will be conducted in Iteration 1, Phase 2, and subtasks (2)-(3) conducted Iteration 2, Phase 3.

Task 6.5 Validation of the Pillar III use cases (M30 – M36) (NGI, ETHZ, INGV, UMA)

We will perform drill runs of representative cases in production-like environments. We will assess the use cases' performance, reliability and added value using the metrics defined in task 6.1 with respect to the alternatives existing prior to the project. At this point we will compare our tsunami wave height and shake estimates with that of current state-of-the-art processes. We will also elaborate on the costs incurred by the runs and their viability as HPC services.

Task 6.6 Integration of results in workflow repository and HPCWaaS interface (M30 – M36) (BSC, UMA, NGI) Developments at workflow level will be compartmentalized as much as possible so that other users can integrate them into their existing workflows. We will commit to standardising the code styling and documentation to the HPCWaaS requirements so that components can be easily used in other areas of the natural hazards' community or even in other communities which can have similar needs (e.g. access to urgent simulations or management of embarrassingly parallel scenario runs).

Task 6.7 Access policy for urgent computing at HPC infrastructures (M6 – M30) (PSNC, BSC)

We will develop a protocol for urgent access to Tier-1 and Tier-0 systems. We will develop automation rules, evaluate the cost and benefit of different scenarios and make recommendations for the governing bodies of such systems regarding the implementation. This should include not only availability and QoS of the infrastructure but also early risk estimates related to the hazard, which should be standardized across different hazard types.

Deliverables

D6.1 Requirements on the eFlows4HPC software stack from Pillar III and evaluation metrics (BSC, R, PU) [M4]. Summary of conclusions on the pillar III requirements.
D6.2 Description of the use cases for Pillar III (INGV, R, PU) [M6]. Compendium of datasets and models employed for development and validation of Pillar III workflows.
D6.3 Iteration 1 workflows for urgent computing of natural hazards (UMA, OTHER, PU) [M18]. Task-based version of UCIS4EQ and PTF workflows.
D6.4 Iteration 2 workflows for urgent computing of natural hazards (BSC, OTHER, PU) [M30]. Final releases of the UCIS4EQ and PTF workflows.
D6.5 Database of Earth models (ETHZ, OTHER, PU) [M30]. Release of Earth models obtained for the use-case regions, together with associated metadata.
D6.6 Protocol for urgent HPC (PSNC, R, PU) [M36]. Protocol definition and governance recommendations regarding urgent HPC.

WP number	7				**Duration**						M01 – M36						
WP Title	Outreach to Communities and Exploitation																
Participant number	1	2	3	4	**5**	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	BSC	CIMNE	FZJ	UPV	**ATOS**	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	12	2	2	2	**15**	1	2	1	1	1	1	1	1	1	1	1	45

The goal of this WP is to develop strategies to a) exploit the outcomes of the project, and b) educate the relevant communities in relationship with the three Pillars and the main technological components of the project.

Objectives

- To train the communities and to identify outreach activities addressed to students;
- To organize three workshops for end-user and communities;
- To identify the exploitation potential of produced results by all project partners, to schedule and monitor the exploitation activities undertaken by the project partners;

Description of work

This WP aims at a performing the corresponding activities to reach the potential user communities of the project results as well as looking at their exploitation opportunities. The WP is organized in tasks that aim at external training and workshops, focusing mainly on the communities linked to the project Pillars (manufacturing, climate and urgent computing (tasks 7.1 and 7.2). The WP will also perform the exploitation activities, including a market analysis and the development of an exploitation strategy (task 7.3).

Task 7.1 Training (M1 – M36) (BSC, all)

Internal training courses will be organized during the project addressed to the users, developers and also related communities. Initially, the idea is to organize a first training course at the beginning of the project, a hackathon across all Pillars and a total of three additional courses (one by each pillar), ideally linked with an event of the community. On a one-to-one basis, it will be considered to open them to other communities, users and developers. A detailed training plan and its needs will be defined in D7.1 that will provide further details about the training courses, timetable, materials, etc.

In terms of outreach, this task will identify outreach activities addressed to students, such as school visits, open days, participation in summer schools, presentations to schools, etc. and participation in science fairs addressed to the general public, whenever applicable. The rest of partners will contribute to this task with their own outreach activities.

Task 7.2 Organization of community workshops (M30 – M36) (BSC, all)

During the second part of the project, in collaboration with the FocusCoE CSA, and with the specific aim to engage the HPC and related research community, three joint workshops with HPC CoEs such as, for example, ChEESE, ESiWACE, EXCELLERAT (and potentially others that are relevant to the Pillars) as potential with eFlows4HPC end-users will be organized. The idea is to co-locate these workshops within an existing conference series, such as the EuroHPC Summit Week conference series. These joint workshops aim to find synergies between eFlows4HPC and the potential communities behind the CoEs and to communicate the developed pillar workflows and the workflow methodologies to different communities. The workshop format plans to include presentations from the

invited CoEs and eFlows4HPC followed by a discussion panel. These joint workshops with HPC CoEs can also be organized with relevant FET HPC projects, or any other relevant EU initiative for eFlows4HPC.

Task 7.3 Exploitation (M01 – M36) (Atos, all)

This task will be devoted to the establishment of the project's exploitation plans. Individual partners will exploit the project's results in-line with their business and/or research strategies and on the basis of modalities such as enhancement of their existing products, use of eFlows4HPC outcomes for improving their software development processes, provision of technology transfer and consulting service. The task will begin with the analysis and exploration of relevant eFlows4HPC applicable markets, which will be turned into an in-depth market analysis. The execution of an in-depth **market analysis** (D7.2) will be presenting the competitive landscape with regard to solutions coming out of the three Pillars' use cases and for the main technological components, i.e. software stack, workflow services, etc. This will allow to carve a niche positioning for eFlows4HPC outcomes and build the value proposition for the adoption of the project's key exploitable assets by the community. On this basis, a comprehensive **exploitation strategy** (D7.3) will be generated, laying out exploitation scenarios for the sustainable uptake of the different components and an action plan to realize this. In addition, the strategy will incorporate the individual exploitation perspectives of the different partners. The first version of the strategy will be delivered at M15; it will be maintained as a live document throughout the project and it will be updated with specific post-project exploitation actions at its final version on M36 (D7.5). All industrial and research partners will contribute to the task because all partners have exploitation interest based upon the technological software stack WPs identified in section 2.2.3. The Pillars use case partners have also major interest and thus contribute in the exploitation planning.

Deliverables

D7.1 Training Plan (BSC, R, PU) [M4]. This deliverable will define the objectives of the project's training activities, the initial plans for organization of the training activities, as well as the materials that will be provided.

D7.2 Market analysis (Atos, R, CO) [M15]. This deliverable will be the first version of the exploitation plans and will provide the market analysis.

D7.3 Exploitation strategy (Atos, R, CO) [M24]. This deliverable will enrich the exploitation plans with the exploitation strategy incorporating the individual exploitation perspectives of the different partners and documenting scheduled exploitation activities of the project results.

D7.4 Report of the organization of community workshops (BSC, R, PU) [M36]. This deliverable will report on the community workshops organized during the lifetime of the project.

D7.5 Exploitation report **(Atos, R, CO) [M36].** This deliverable will update the document with activities undertaken during the project.

WP number	8				**Duration**									M01 – M36			
WP Title	Dissemination and Communication																
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	**TOTAL**
Participant short name	**BSC**	CIMNE	FZJ	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	
PMs per participant	**9**	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	**26**

Objectives

- To build up an eFlows4HPC brand, as well as disseminate the project's ideas and results via the public website and dedicated and partners' social media channels, participation in conferences and presenting papers in scientific journals
- To continue building a dynamic community of researchers and application designers who engage with the project via the public website and LinkedIn account.
- To communicate the benefits of eFlows4HPC's research to industry stakeholders and the wider public via targeted events, trainings and project dissemination materials.

Description of work

The dissemination and communication team will work closely with other WPs to obtain up-to-date information on the achievements of the project in order to disseminate them to the appropriate audiences. Specific plans will be produced related to Dissemination and Communication plan (D8.1) and their corresponding updates in later reports

(D8.2, 8.3 and D8.4). WP8 activities will liaise closely with other WPs in order to provide up-to-date information on the achievements of the project, and disseminate them to the appropriate audiences.

Task 8.1: Dissemination and Communication Strategy (M1-M36) (BSC, all)

An effective communication strategy is crucial to ensure the coherence and effectiveness of the project activities. It will include activities which will aim both to raise awareness of the project and its aims to engage key stakeholders, an aspect which is particularly important given that their input will be required during the iterative development process.

In this task, the communication team will define the objectives of the communication plan, identifying target audiences and setting out appropriate channels and messages to communicate with them, before defining activities and tools to achieve these objectives. The outcome of this task will be the Communication Plan (D8.1). This document will be updated during the project's lifetime as required. All activities will aim to have a direct impact on policymakers, companies and also to society in general. The impact of all activities will be carefully considered and analysed afterwards.

Task 8.2: Dissemination materials (M1-M36) (BSC, all)

The communication team will carry out the activities detailed in the communication plan. These activities will include: Identification of the relevant communities which make up eFlows4HPC target audiences and relevant communication messages and channels for each; creation and maintenance of a public website which will provide clear information about the project, allow dissemination of project documents and results, keep readers updated on news and events and provide a hub for the online community following the project; production of online and printed materials to disseminate information about the project. These might include posters, flyers, factsheets, booklets, presentations and audio-visual materials such as videos or animated gifs with the acknowledgement of the EC funding sources will be included in all communication materials; a dedicated Twitter account to help sustain a digital community and keep followers updated on project progress; identification of targeted events where the project should be presented and peer-reviewed journals to which papers can be submitted; and creation and distribution of press releases.

Task 8.3: Organization or participation in events and outreach (M1-M36) (BSC, all)

This task will focus on collecting all events for the consortium to attend and disseminate eFlows4HPC's results, as well as participating in existing exhibitions from partners in ISC and SC conferences. It will also be considered to host workshops for each pillar; a detailed plan will be added in D8.1. These workshops will be organized in locations determined by the organizing partner, it will have a registration fee, and a professional congress organizer will be hired for the logistical support. Its aim is to train the defined sector Pillars as well as to give visibility to the project.

Deliverables

D8.1: Dissemination and Communication Plan (BSC, R, PU) [M3] This deliverable will set out the dissemination and communication strategy and the activities to be undertaken to achieve this. Results of the dissemination work will be reported in the periodic and final reports.

D8.2: First Dissemination and Communication Report (BSC, R, PU) [M12] This deliverable will report on the dissemination and communication activities of the project done during the first year.

D8.3: Second Dissemination and Communication Report (BSC, R, PU) [M24] This deliverable will report on the dissemination and communication activities of the project done during the second year.

D8.4: Final Dissemination and Communication Report (BSC, R, PU) [M36] This deliverable will report on the final dissemination and communication activities of the complete project.

WP number	9				**Duration**						M01 – M36						
WP Title	Project Management																
Participant number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**TOTAL**
Participant short name	**BSC**	CIMNE	FZJ	UPV	ATOS	DtoK	CMCC	INRIA	SISSA	PSNC	UMA	INGV	AWI	ETHZ	Siemens	NGI	

PMs per participant	**20**	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	25

Objectives

- To ensure a clear and effective liaison with the EC, and communication between partners, detecting management and technical issues as early as possible and bring them to resolution.
- Continuous monitoring to deliver objectives within the time and budget constraints of the project.
- To establish and enforce effective management and quality procedures to produce high quality deliverables.
- To provide efficient operational management support including: administrative and financial planning, reporting to the EC and management of project legal aspects including project-related contracts and IPR, and management of day-to-day operational and technical progress.
- Ensure compliance with the "Ethics Requirements" along the project.
- Deliver the project data management plan

Description of work

This WP will ensure the correct execution of the Project. This will be managed and directed according to the work plan authorised by the European Commission, and the management will make sure that the requirements set by the Commission concerning reporting, information, etc. will be fulfilled. The project manager will coordinate the project from the global point of view considering all organisational, legal, and financial aspects, including the advisory board management, Data Management Plan and ethics requirements.

Task 9.1: Legal, Financial and Administrative Management (M1-M36) (BSC, all).

This task will establish the corresponding procedures, tools and methodologies to enable a correct project management, including legal, administrative and financial. It will also coordinate the timely production of deliverables, organize the kick-off meeting, technical meetings and reviews, and organize and manage audits requested by the commission. On a 6-month basis, the project coordinator will monitor resources usage, producing internal use of resources reports to ensure the project resources expenditure is in track with the work progress. At the end of each period, the Project Coordinator will prepare a technical progress report describing the project activities performed during the period against plan, based on the Guidelines provided by the European Commission for H2020 reporting. Finally, at the end of the project, a Final Report will be prepared based on the Guidelines provided by the European Commission for H2020, including: a final publishable summary of the work completed to date covering results, the conclusions and socio-economic impact of the project, a chapter on awareness and wider societal implications as well as a report on the distribution of the Community financial contribution.

Task 9.2: Project coordination (M1-M36) (BSC, all)

This task will perform the technical coordination of the project, by means of monitoring the progress of the WPs, technical coordination of the meetings, appointing reviewers to assess the quality of the deliverables before their delivery to the EC, and solving technical conflicts.

Task 9.3: Internal communication, Quality and Risk management (M1-M36) (BSC, all)

This task will determine the appropriate strategy to ensure clear communication channels between all partners in order to facilitate the exchange of critical project documentation and news, and to encourage participation in the decision-making process. The task will define and maintain internal collaborative tools for sharing documentation and communicating work status. One of the outcomes of this task will be an intranet (shared workspace in the website) and a series of distribution lists. This task will also define and implement the appropriate quality assurance processes to ensure an accurate documentation, reporting and justification of the work being carried out. A process will be developed (and tools if required) to ensure that the deliverables are reviewed by a broad spectrum of individuals against a well-defined set of criteria. Moreover, it will determine the minimum level of quality required for presentation of the official outcomes of the project to the EC. The high-level principles guiding these procedures will be agreed to at the start of the project at the Kick-off Meeting.

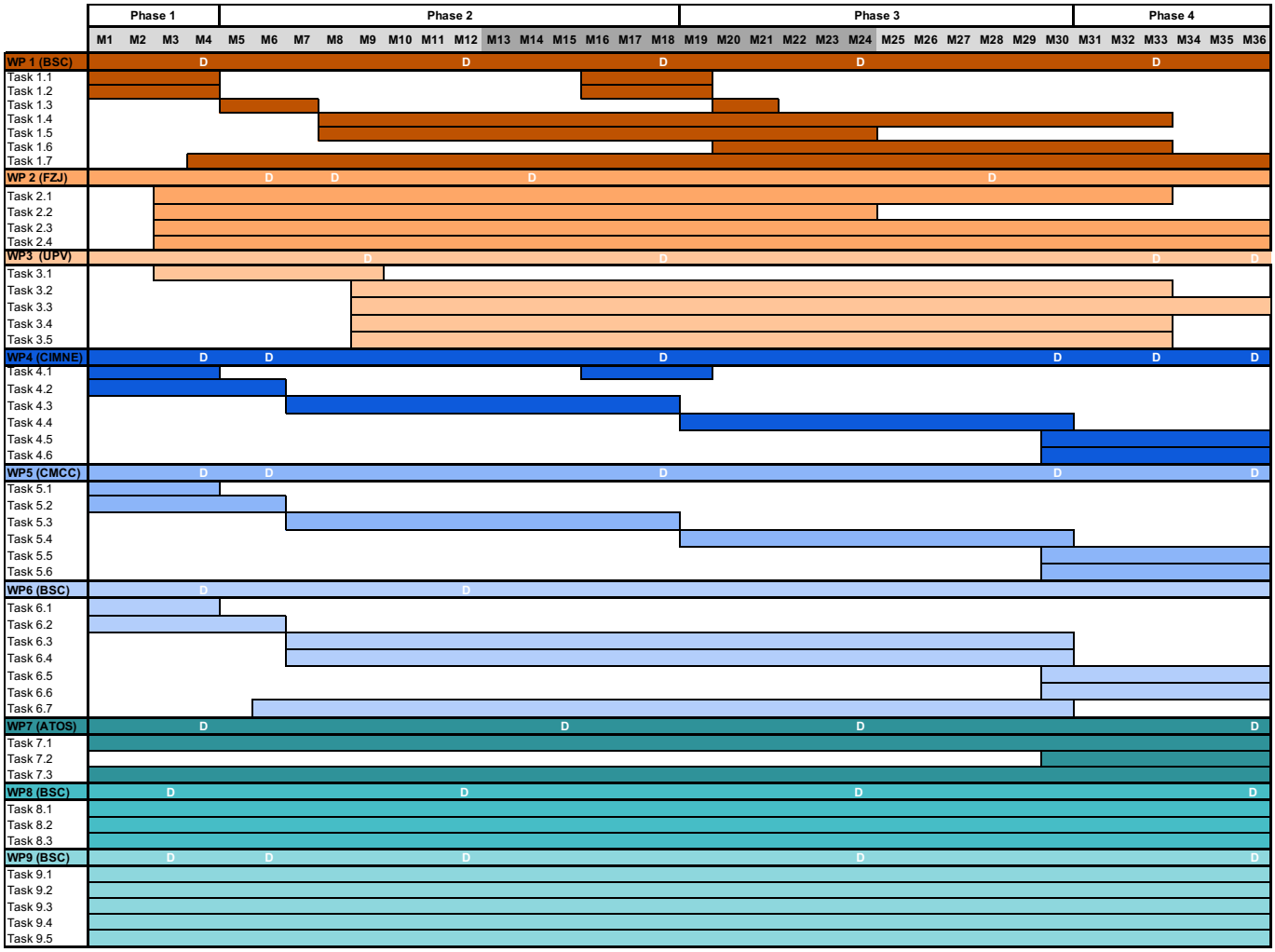
Task 9.4: Data Management Plan (DMP) (M1-M36) (BSC, all). The DMP will describe the life cycle for all datasets that will be collected, processed or generated by the research project, providing an analysis of the main elements of the data management policy that will be used by the applications with regard to all the datasets that will be generated by the project, according to the Guidelines on Data Management in H2020. The DMP will evolve during the project with new versions released at every project stage.

Task 9.5: Advisory Board Management (M1-M36) (BSC, all). During the first months of the project, the final composition of the Advisory Board will be defined, starting with the current members engaged as detailed in Annex B. The composition will take into account the different stakeholders in the value chain, so they are represented there. Collaboration agreements will be put in place also. This task then will manage all the communication and relationship with the Advisory Board Members during the project lifecycle.

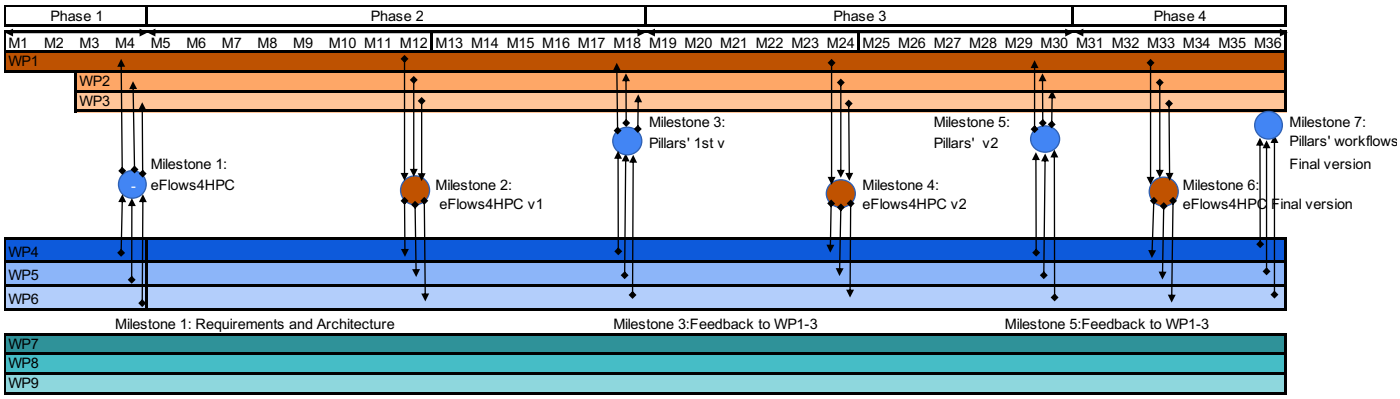
Deliverables

D9.1 Project management plan and quality guidelines (BSC, R, CO) [M3]. Document (handbook) that summarizes the main project procedures (contractual framework, project governance, project monitoring, internal communication, templates available, etc.) as well as the standards and quality procedures to be used throughout the project lifecycle.
D9.2 Data Management Plan (DMP) (BSC, ORPD, PU) [M6]. Document outlining how data will be managed during the project from internal and external point of view. The DMP will include a table specifying how the data will be exploited, shared for verification and reuse. Updates to this report will be provided in **M12, M24** and **M36.**

3.2 eFlows4HPC Gantt Chart



Milestones and relation to WPs



3.3 List of deliverables

Del No	Deliverable title	Lead Partic.		Deli. Date	Diss. Lev	Nature
D1.1	Requirements, metrics and architecture design	1	BSC	M4	PU	R
D1.2	eFlows4HPC interfaces and first software stack release	6	DtoK	M12	PU	R+O
D1.3	Revision of Requirements and architecture design (BSC, R, PU) [M18]	1	BSC	M18	PU	R
D1.4	eFlows4HPC interfaces and second software stack release	5	ATOS	M24	PU	R+O
D1.5	eFlows4HPC interfaces and final software stack release	1	BSC	M33	PU	R+O
D2.1	Data Catalogue	3	FZJ	M6	PU	R+O
D2.2	Technology Evaluation, Containerization and Optimization Strategy	5	ATOS	M8	PU	R
D2.3	First version of Data Logistics	3	FZJ	M14	PU	R+O
D2.4	Report on Implementing Containerization and Optimization Strategy	5	ATOS	M28	PU	R+O
D2.5	Final Report on Data Logistics Implementation	3	FZJ	M36	PU	R+O
D3.1	Application bottlenecks and optimization opportunities on heterogeneous components	4	UPV	M9	PU	R
D3.2	Initial draft of optimized kernels for heterogeneous components	4	UPV	M18	PU	R
D3.3	Initial draft of optimized kernels for EPI	1	BSC	M18	PU	R
D3.4	Optimized kernels for heterogeneous components	4	UPV	M33	PU	R
D3.5	Optimized data management with new storage technologies	1	BSC	M33	PU	R
D3.6	Optimized kernels for EPI	1	BSC	M36	PU	R
D4.1	Requirements on the eFlows4HPC software stack from Pillar I and evaluation metrics	2	CIMNE	M4	PU	R
D4.2	Design of the Pillar I use cases	2	Siemens	M6	PU	R
D4.3	ROM Tools Release	2	SISSA	M18	PU	O
D4.4	Demo ROM	2	CIMNE	M30	PU	O
D4.5	Release of HPCWaaS integrated solver stack	2	BSC	M33	PU	O
D4.6	Validation of the Pillar I use cases	2	Siemens	M36	PU	R
D5.1	Requirements on the eFlows4HPC software stack from Pillar II and evaluation metrics.	13	AWI	M4	PU	R
D5.2	Design of the Pillar II use cases	1	BSC	M6	PU	R
D5.3	Pillar II - Iteration 1 Software Release	7	CMCC	M18	PU	O
D5.4	Pillar II - Iteration 2 Software Release	7	CMCC	M30	PU	O
D5.5	Pillar II Validation of requirements	13	AWI	M36	PU	R
D6.1	Requirements on the eFlows4HPC software stack from Pillar III and evaluation metrics	1	BSC	M4	PU	R

D6.2	Development plan for natural hazards' urgent computing workflows	12	INGV	M6	PU	R
D6.3	Description and demonstration of technologies released to HPCWaaS	11	UMA	M18	PU	O
D6.4	Technological improvement in natural hazards' urgent computing workflows	1	BSC	M30	PU	O
D6.5	Database of Earth models	14	ETHZ	M30	PU	O
D6.6	Protocol for urgent HPC	10	PSNC	M36	PU	R
D7.1	Training Plan	1	BSC	M4	PU	R
D7.2	Market analysis	5	ATOS	M15	CO	R
D7.3	Exploitation strategy	5	ATOS	M24	CO	R
D7.4	Report of the organization of community workshops	1	BSC	M36	PU	R
D7.5	Exploitation report	5	ATOS	M36	CO	R
D8.1	Dissemination and Exploitation Plan	1	BSC	M3	PU	R
D8.2	First Dissemination and Communication Report	1	BSC	M12	PU	R
D8.3	Second Dissemination and Communication Report	1	BSC	M24	PU	R
D8.4	Final Dissemination and Communication Report	1	BSC	M36	PU	R
D9.1	Project management plan and quality guidelines	1	BSC	M3	CO	R
D9.2	Data Management Plan	1	BSC	M6	PU	ORPD

3.4 Management structure, milestones and procedures

The management structure for eFlows4HPC is derived from successful models used by the partners in previous H2020 projects, considering the specific needs of a project that aims to deploy a complex integrated system in a short timeframe as well as the contractual restrictions of the Horizon 2020 Framework Programme. eFlows4HPC is an innovation-focused project consisting of 16 partners (4 universities, 9 research centres and 3 industries,) that contribute to the project with complementary expertise. It is important to note that each project partner has been involved in previous European and National Projects.

3.4.1 Project Management Principles and Structure

The goal of eFlows4HPC management is to lead the project to achieve its goals, providing a light-weight management service able to ensure: (1) flexible and open-intended dialog among partners in all key strategic and technological issues; (2) fast and effective decision-making procedure on technical and organizational issues; and (3) compliance with EU contractual requirements.
The project organizational structure includes the key components illustrated in the figure:

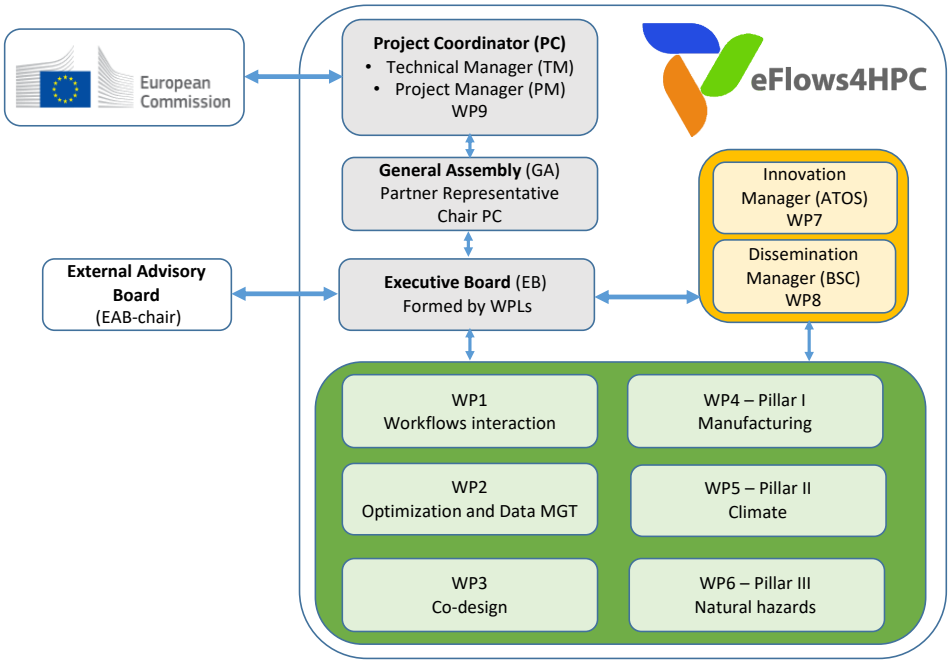


Figure 9: eFlows4HPC management structure

The **Project Coordinator (PC)**, Dr. Rosa M. Badia (BSC), has the responsibility on the managerial and scientific aspects of the project. She is also the **Technical Manager (TM)**, has the overall responsibility for the project progress, and guarantees that the scientific and technical objectives of the project are met. The TM works with the Work Package Leaders (WPL) to identify issues and propose suitable corrective actions (e.g., temporary resource reallocation, taskforce creation, etc.) that might require approval by the Executive Board (EB). The TM is supported by the Project Manager (PM), collaborating closely to provide clear and accurate Periodic Reports.
The **Project Manager (PM)**, Vanessa Fernandez (BSC), is responsible for the day-to-day execution of the project and will ensure the timely delivery of project objectives and deliverables by continuously monitoring how closely the project progress is following the plan. The PM is also responsible for the administrative management of the project that includes the provisioning of Periodic Reports and Financial Statements to ensure a timely distribution of the budget to the beneficiaries according to the grant agreement.
The **General Assembly (GA)**, chaired by the PC, is the decision-making body and is formed of a delegate from each partner. Each partner will have one vote, with the vote of the chairperson deciding in case of a tie. The GA will provide a forum for the discussion of administrative and strategic management issues linked to the project. It will decide on approving major modifications to project plans, allocated efforts, and budget issues. Face-to-face 2-day or 3-day meetings will occur for the kick-off of the project (M1), for the consolidation of requirements and architecture definition (M3), during the first software stack integration round (M9), and before the first pillar release round (M15), to speed up integration. Similarly, during the third and four Phases, there will be a meeting to prepare the second release of the software stack (M22), to prepare the second release of the pillar workflows (M28), and a meeting to revise the workflows validation (M34). Additional workshops, bilateral meetings, and teleconferences will be set on demand to address any challenge hindering the progress of the project.
The **Executive Board (EB)** is the main day-to-day decision-making team, it is chaired by the TM and it includes the WPLs. The EB is an executive body that reviews the project progress on a regular basis; it has ample powers to make decisions on daily implementation issues and it is responsible for resource allocation, the review/approval of the Periodic Reports and Deliverables, the preparation of project reviews and the coordination of exploitation plans.
The EB will hold monthly conference calls and will make decisions by consensus. In case the EB cannot obtain consensus with respect to an issue, the issue is forwarded to the GA and brought to a vote if required. The EB will hold regular monthly teleconferences to evaluate progress, assess risks, and take any decision needed to timely meet the project goals.
The **Work Package Leaders (WPLs)** are responsible for the scientific and technical work of their respective WPs. This includes planning and control of all activities and the collection of the contributions from partners participating in the WP. Technical discussions and work will be led by WPLs, while inter-WP issues will be solved by the EB.
The **External Advisory Board (EAB)** is a panel of independent experts, including representatives of academia, companies, standardisation bodies, and end-users outside of the consortium, and who are therefore not involved in the day-to-day project work. The final configuration of the EAB will be selected after the kick-off meeting to prevent possible conflicts and within the first three months of the project by the EAB Chair, although some members have

been already invited and their bios are given in Annex B. The EAB members will provide comments and recommendations regarding requirements, objectives, and development, as well as exploitation and dissemination activities. This will ensure the maximum impact of the eFlows4HPC project and the future exploitation of the project results. Such recommendations will be discussed and processed in GA meetings and included in the PPRs and in the first and final period reports. The involvement of EAB experts in exploitation activities and the access rights to the foreground IP will be regulated by the Consortium Agreement, which will be signed by consortium partners prior to the project start date.

The Innovation Manager (IM) facilitates the migration of the exploitable project foreground into products or IPs suitable for the reference markets. The IM will act in close coordination with the PC, and the EAB, and will advise the GA and PC on responses to changes in the industry scenario and roadmaps. The IM will also lead task 7.3 Exploitation, in order to: (1) coordinate exploitation strategies amongst users; (2) monitor use and dissemination of Knowledge; (3) collect IPR information on project foreground, advising the owners on protection means, agreeing with their owner the access conditions for the project and then for the usage, keeping track of licensing within the project, moderating potential IPR conflicts; (4) assist dissemination and exploitation activities from the IPR perspective; (5) ensure a consistent approach in dissemination activities by reviewing materials before their release to remove sensitive information and protect IPR through appropriate measures (copyright, patents, etc.); and (6) track/monitor the TRL during the project execution. The IM is Atos. The IM will closely cooperate with the Dissemination Manager (DM) to raise awareness of the project, to maximise its visibility and to ease the uptake of its outputs outside the Consortium. The DM will be BSC.

Dedicated services and procedures and infrastructures will be defined and adopted to favour information flow and communication. A project web server will serve as a repository for all the working documents. A public part will publicize the description of the project objectives, partner roles, public deliverables, and news. An additional private part will instead be devoted to internal documents. The source code will be released in public repositories, like github. Internal communications will be made via email and communication tools (i.e. Slack), project meetings, and telcos. An email list will be set up by the coordinator, and its contents will be backed up. Management reports will be prepared in a structured fashion for periodic reporting to the EC. The main content will be the progress reports and management reports from the WPLs, integrated by the TM and the PM. Deliverables will be available for partners in the private area of the web server.

Consortium Agreement. The Consortium Partners will conclude a Consortium Agreement (CA) which will be signed before the project starts (before the signature of Grant Agreement). CA will include the internal Management guidelines, consortium organization, fund distribution, management of background and foreground including IPR management during and after the project, and transfer of the eFlows4HPC project results.

Conflict resolution. The frequent interaction needed in the project for tools and service integration requires the use of a lightweight decision process. Decisions will be taken during project meetings, telcos or by email, and will be backed up with appropriate written confirmations in the form of emails, document sections, or formal meeting minutes, which all partners will be able to review for final approval. Agreement will be reached whenever possible by consensus. If this fails, the TM will prepare a focused meeting or telco with all partners involved, possibly with proposed resolutions. Any event that may jeopardize the overall completion date of the project should be reported immediately to the TM. The TM will call an emergency EB meeting or teleconference as required. Each party involved in the issue must present a short document describing its respective understanding of the conflict that includes at least one proposed solution. The EB reviews the conflict documents and following the procedures of the EB, each member votes on one of the proposed solutions. The solution receiving the simple majority is implemented with the TM, as chairperson, casting the tie-breaking vote as necessary.

3.4.2 List of milestones

Milestone No.	Milestone name	Related WP(s) (main underlined)	Due date (in month)	Means of verification
MS1	eFlows4HPC requirements and architecture	WP1, WP4, WP5, WP6, WP7, WP8	M4	Requirements and architecture defined. Definition of metrics. All available in deliverables D1.1, D4.1, D5.1 and D6.1. Definition of the training plan (D7.1) and of the communication and dissemination plan (D8.1).
MS2	eFlows4HPC software stack version 1	WP1, WP2, WP3, WP4, WP5, WP6	M12	First version of the eFlows4HPC software stack available in public project repository (D1.2). Project data catalogue available (D2.1).

				Deployment strategy available (D2.2) Key application kernels and hardware-related bottlenecks identified. Detailed performance and energy efficiency analysis on target kernels (D3.1). Design of Pillars' use cases available (D4.2, D5.2, D6.2)
MS3	First version Pillars. Revision of requirements and architecture	WP1, WP4, WP5, WP6, WP7	M18	First version of Pillars' workflows integrated with eFlows4HPC software stack and available in the public repository of code (D4.3, D5.3, D6.3). Revision of the requirements on the software and architecture defined and available (D1.3). First version of data logistics available (D2.3). Optimized kernels available (first version, D3.2, D3.3). Hackathon across the three Pillars organized. First version of the exploitation plan with market analysis available (D7.3).
MS4	eFlows4HPC version 2	WP1, WP2, WP3, WP7	M24	Second version of the eFlows4HPC software stack available in public project repository (D1.4). Exploitation plan enriched with the exploitation strategies and scheduled exploitation activities (D7.4)
MS5	Second Version Pillars	WP2, WP4, WP5, WP6	M30	Second version of Pillars' workflows integrated with eFlows4HPC software stack and available in the public repository of code (D4.4, D5.4, D6.4) Containerization and Optimization Strategy available (D2.4). Data logistics service available (D2.5).
MS6	eFlows4HPC final version	WP1, WP2, WP3	M33	Final version of the eFlows4HPC software stack available in public project repository (D1.5). Optimized kernels for new heterogeneous components and storage devices available. Detailed performance and energy efficiency analysis and comparison on target kernels and with specialized hardware components (D3.4, D3.5). Storage optimizations integrated in workflows (D3.6).
MS7	Pillars workflows final version available in repository	WP4, WP5, WP6	MS36	Outcome of the Pillar results available to the public (D4.6, D5.5). Organization of the Community Workshops. Report of the exploitation activities undertaken by the project. Workflows available in workflow registry. Reusable models available in model repository. Libraries catalogue populated. Protocol for Urgent Computing defined (D6.5).

3.4.3 Risks management and quality control.

A risk management and quality assurance process will be defined (T9.3) to ensure accurate documentation, reporting, and justification of the work. An internal peer-review process will be set up, including well-defined criteria, to assure the project deliverables meet the quality standards. The high-level principles guiding these procedures will be agreed at the start of the project. Risks that could affect the full accomplishment of the objectives may arise due to the complex activities of the project. These have been identified in advance, and mitigation measures have been arranged for each case (see table below). The management structure is prepared to deal with these risks by applying risk management, including monitoring on the severity of the risk and the follow up of its evolution and mitigation, following the communication flow described.

Description of risk (level of likelihood: Low/Medium/High)	WP(s) involved	Proposed risk-mitigation measures
The requirements' analysis and architecture definition phase does not converge after the first four months. (Low)	WP1, WP5, WP6, WP7	The project coordination will make sure this initial phase starts with all necessary efforts to guarantee convergence.
Issues on the integration of the different components of the workflows' software stack (Low)	WP1	The project methodology has been carefully designed with phases to analyse the requirements, architecture and interfaces that makes this risk very low. In the case of occurrence, the continuous development and integration methodology will enable to detect possible issues and to establish corrective measures.
Technology and platforms "impedance mismatch". eFlows4HPC will comprise of (mostly) existing technologies and services currently used in specific environments. In the project, some of the technologies will be put on new platforms, it is hard to see how well this porting will work out. (Low)	WP2	Technology evaluation must include information about current running environments and sought after the reasonable behind it.
Computation over data paradigms clash. Current approaches favor one paradigm: either the computation is moved to data or other way round. The envisioned workflows will include processing of both types, requiring architectural decision which paradigm should be followed or how both can be reconciled. (Medium)	WP2	The optimization step will collect performance metrics giving hints on wrong selection of a paradigm and the resulting penalties. This information can be used to reimplement parts of the workflows.
Delayed delivery of applications results in late identification of kernels (Low)	WP3	Suitable additional benchmarks could be added to the set of kernels to exercise specific eFlows4HPC technologies.
Problems of integration and performance of target kernels into the Pillars' applications (Med)	WP3	Redesigning the affected parts of the Pillars' applications that prevent achieving the expected benefits.
Physics is not reducible by ROM approaches or predictive model is not available for specific features of the solution (Med)	WP4	Combine FEM-Based ROM for the well-behaved components of the solution with simplified "heuristic" models based on AI to cover the problematic feature.
The model is too expensive even after hyper-reduction (Low)	WP4	Adopt a modular approach in decomposing the physics and eventually adopt ML black-box approaches for specific features in order to reduce the overall computational cost.
User requirements list is not complete or key requirements are missing or not adequately captured (Low).	WP1, WP4, WP5, WP6	Several partners in the consortium are strongly involved in the relevant domains/communities as well as key initiatives and scientific activities related to the pillar use cases.
Climate use cases not being released on time due to complexity of integration of new modules (features extraction, AI-assisted support for ensemble simulations) into legacy software and operational workflows (Medium)	WP5	Partners in the consortium have a strong knowledge of legacy software (i.e. ESM models) and interfaces as well as HPC production systems and operational workflows. Early prototypes will be developed on well-isolated core components during Iteration 1, while full integration will be addressed in Iteration 2.
eFlows4HPC solutions are not pervasively adopted/exploited by the different target communities (Medium)	WP4, WP5, WP6, WP7,	The activities and solutions will be properly disseminated and demonstrated in relevant domain-specific meetings and workshops with a wide participation of community representatives as well as in internal and external

	WP8	trainings. Partners in the consortium are also actively involved in the EXCELLERAT, ESiWACE2, and ChEESE CoEs as well as other related EU-projects/initiatives (i.e. EXDCI, PRIMAVERA).
Missing effective transfer to end-users due to underestimation of uncertainty within operation time windows (Medium)	WP6	There is a trade-off between computation time and ensemble size. Difficulties may emerge in simulating sufficiently large ensembles in the operational time frame of end-users (e.g. 2 hr for ARISTOTLE). This risk can be mitigated by increasing computational capabilities, implementing further optimizations in the workflows and in the ensemble management (subsequent refinements, advanced sampling techniques, etc.), potentially mixing physically-based simulations with alternative (and faster) simulation strategies (e.g. empirical models, ROM-like emulators/meta-models).
Exploitation strategy failed to the market opportunities (Low)	WP7	This will hardly happen because a) a market analysis will be done at an early stage of the project, and b) the exploitation strategy will be led by an experienced and professional team and will stick on the market opportunities explained in the analysis. Nevertheless, market opportunities may be identified by other partners in the domain, or later in the project, but within its lifecycle. Should this happen, the exploitation strategy leaders will evaluate the opportunities, and try to re-formulate / adapt the exploitation strategy accordingly in its final iteration to facilitate the exploitation of these opportunities.
Low interest and participation in the dissemination activities (Low)	WP8	All the activities will be actively promoted using all the instruments of eFlows4HPC partners and their participation in relevant communities and initiatives. If lower attendance occurs, efforts will be increased.
Lack of required know-how, possibly due to the departure of a key member (Medium)	WP9	Risk is minimized since the involved partners' teams have experienced personnel. Partners will foster technical excellence to ensure competence. If needed alternative personnel will be allocated
Delays caused by not meeting the task deadlines. (Medium)	WP9	Partners involved in delayed tasks will allocate additional resources to meet planned deadlines. Agile project management and frequent calls will mitigate this risk.
Partner leaves the Consortium due to unexpected reasons (Low)	All WPs	eFlows4HPC partners have a very well-known activity in European Projects so there this possibility is very unlikely. In such a case, the GA will check if the uncovered activities can be covered by other partners or a new one must be added.
Low exploitation potential of results at the end of the project (Low)	All WPs	The Consortium includes recognised EU industrial leaders, community members, and some of the best universities and research centres in Europe. The industrially and community driven research that is part of the project strategy will prevent this risk from arising.

3.5 Consortium as a whole

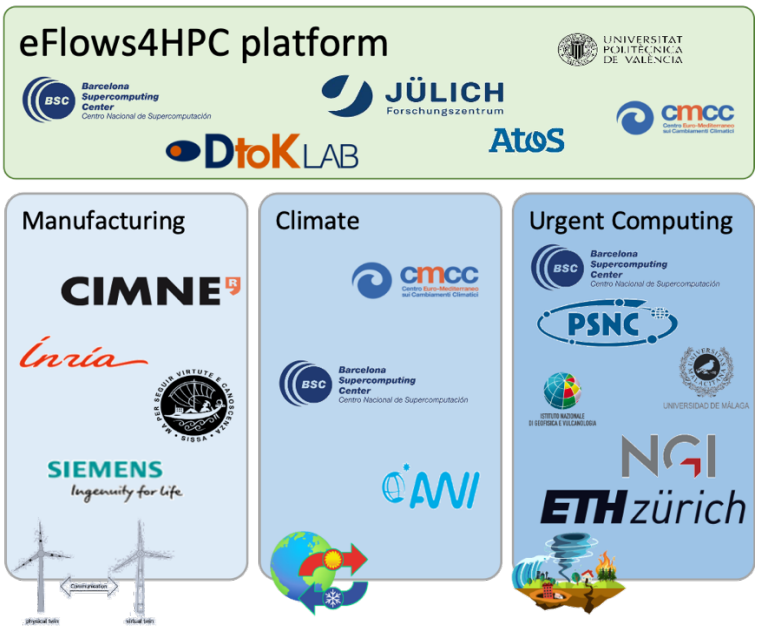


Figure 10: Consortium composition

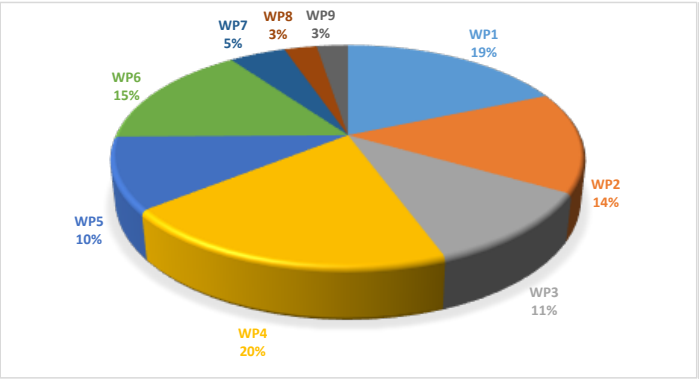
The eFlows4HPC consortium comprises 16 partners from 7 different countries. The consortium has been carefully designed to obtain a strong team with complementary expertise in technical aspects: supercomputing and acceleration (BSC, FZJ, PSNC, UPV, Atos), workflow management and orchestration (BSC, Atos, CMCC), machine learning (FZJ, UPV, Atos, DtoK), big data analytics (DtoK, Atos, CMCC), data management (FZJ), storage (BSC); together with expertise in the pillar workflows' areas: manufacturing, climate, urgent computing. Each of the Pillars is composed of a set of partners with complementary expertise. In the manufacturing pillar, the team members have leading expertise in all of the areas involved: general purpose simulation software (CIMNE), immersed solution processes (CIMNE, INRIA), parallel mesh adaptation (INRIA), Reduced Order Modelling (SISSA), industrial guidance to test cases definitions (Siemens). The climate pillar includes partners with expertise in Computational Sciences applied to climate change (CMCC, BSC), high-performance computing applied to earth system modelling (BSC, CMCC), high resolution and ensemble climate modelling (AWI, CMCC), and workflow tools, machine learning and big data analytics (CMCC). The urgent computing pillar includes partners with expertise on supercomputing (BSC, PSNC), reference institutions for Tsunamis and Earthquakes (INGV, NGI), Tsunami models (UMA), experts in HPC numerical wave propagation codes for full seismic waveform inversion (ETH Zurich), and experts on preparing the policies for Urgent Computing scenarios and preparing the piloting environment for urgent mode (PSNC).
Furthermore, the partners play complementary roles in the project, classified as supercomputing centre (BSC, FZJ, PSNC), technology provider (BSC, FZJ, UPV, Atos, Dtok, CMCC, Inria, SISSA, UMA, AWI, Siemens), workflow provider (BSC, CIMNE, CMCC, INRIA, SISSA, UMA, INGV, AWI, ETH, NGI), data provider (BSC, CMCC, INGV) and end user (BSC, CMCC, INGV, AWI, Siemens) (see table in Annex D) providing players for each of the value chain components.
The participants form a real multidisciplinary team with experts from different application and technological areas, which will guarantee the successful execution of the work program. The necessary set of skills required is not possible to achieve at national level, requiring participation from partners from multiple countries that can provide the necessary complementary expertise and can play the different required roles.
It is also important to remark that most of the partners have previous large experience en EU and national projects and bring substantial prior work into the project. Also, most of the partners have collaborated between them in previous EU projects.
The table in Annex D summarizes the role of the different partners in the project and the area of expertise covered by each of them and their complementarity.

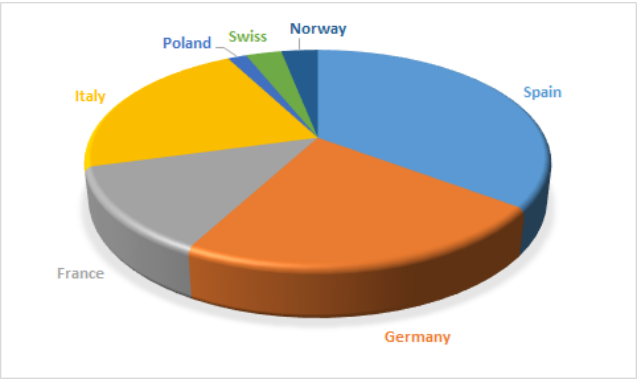
3.6 Resources to be committed

The **total funding** for eFlows4HPC is **7.663.216,25€**, well aligned to the suggestion of this call. The majority of the resources required for the successful completion of the project are human resources, the **16 partners** involved, for 7

different countries, will dedicate a **total of 971 person/months** of effort over the project duration of 36 months. The **eFlows4HPC platform WPs** (WP1, WP2, WP3) keep a **44%** considering they concentrate efforts for the integration and development of the eFlows4HPC platform, comprising the eFlows4HPC software stack and the HPCWaaS methodology. The **Pillars' WPs** (WP4, WP5, WP6) will take **45%** stating the importance and workload associated with the development of the pillar workflows and the validation of the eFlows4HPC platform. The activities related to **Exploitation, Dissemination and Management** stand for **11%** of the effort.

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	Total PMs
BSC	45	36	24	0	36	36	12	9	20	218
CIMNE	0	0	0	81	0	0	2	1	0	84
FZJ	30	47	9	0	0	0	2	1	1	90
UPV	18	8	40	0	0	0	2	2	2	72
ATOS	47	20	0	0	0	0	15	2	2	86
DtoK	28	20	22	0	0	0	1	1	0	72
CMCC	12	12	13	0	32	0	2	1	0	72
INRIA	0	0	0	43	0	0	1	1	0	45
SISSA	0	0	0	43	0	0	1	1	0	45
PSNC	0	0	0	0	0	16	1	1	0	18
UMA	0	0	0	0	0	32	1	1	0	34
INGV	0	0	0	0	0	34	1	1	0	36
AWI	0	0	0	0	34	0	1	1	0	36
ETHZ	0	0	0	0	0	17	1	1	0	19
Siemens	0	0	0	26	0	0	1	1	0	28
NGI	0	0	0	0	0	14	1	1	0	16
TOTAL	**180**	**143**	**108**	**193**	**102**	**149**	**45**	**26**	**25**	**971**





Total direct personnel costs for the project amount to **5.662.489€** (being **74%** of the total project budget). The distribution of the other direct costs (**468.084€**, **6%**) will be mainly dedicated to travel costs (including EAB members) and all Dissemination and Exploitation activities related with this project. These amounts in Euros do not include the 25% overhead.

Other Direct Costs Analysis: below we present the breakdown of "other direct costs" for participants where those costs exceed 15% of the personnel costs.

1 BSC	**Cost**	**Justification / Explanation**
Travel	**54.000€**	Attendance to review meetings, workshop and training events and dissemination activities (43.200€). EAB travel expenses (10.800€).
Other goods and services	**94.334€**	Dissemination material and activities (39.900€). Organization of training events (33.934€). Open access cost for the whole consortium (18.000€). Audit cost (2.500€).
Total	**148.334€**	

10 PSNC	Cost	Justification / Explanation
Travel	**14.800€**	Attendance to review meetings and technical events. Participation in dissemination and exploitation activities.
Other goods and services	**2.000€**	Consumables
Total	**16.800€**	

References

[1] V. G Honavar, et al. Accelerating science: A computing research agenda. arXiv preprint arXiv:1604.02006, 2016.

[2] Asch, M. et al. "Big data and extreme-scale computing: Pathways to convergence-toward a shaping strategy for a future software and data ecosystem for scientific inquiry." The Int. J. of High Performance Comp. Apps 32, no. 4 (2018): 435-479.

[3] Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

[4] Talia, D., "Workflow Systems for Science: Concepts and Tools", ISRN Software Engineering, vol. 2013, pp. 1-15, 2013.

[5] Altintas, I., el al. "Kepler: an extensible system for design and execution of scientific workflows." In Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on, pp. 423-424. IEEE, 2004.

[6] Hull, D. et al. "Taverna: a tool for building and running workflows of services." Nucleic acids research 34, no. suppl_2 (2006): W729-W732.

[7] Afgan, E. et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update." Nucleic acids research 46, no. W1 (2018): W537-W544.

[8] Deelman, E. et al. "Pegasus, a workflow management system for science automation." Future Generation Computer Systems 46 (2015): 17-35.

[9] Fahringer, T. et al. "ASKALON: A grid application development and computing environment." In Proceedings of the 6th IEEE/ACM International Workshop on Grid Computing, pp. 122-131. IEEE Computer Society, 2005.

[10] Manubens-Gil, D. et al. "Seamless management of ensemble climate prediction experiments on HPC platforms." In HPCS, pp. 895-900. 2016.

[11] Manubens-Gil, D., et al.. Assessment report on autosubmit, cylc and ecflow. Tech. rep, 2016.

[12] Lordan, F. et al. "ServiceSs: An interoperable programming framework for the cloud." Journal of grid computing 12, no. 1 (2014): 67-91.

[13] Wilde, M. et al. "Swift: A language for distributed parallel scripting." Parallel Computing 37, no. 9 (2011): 633-652.

[14] https://www.elixir-europe.org

[15] Gropp, W. et al. Using MPI: portable parallel programming with the message-passing interface. Vol. 1. MIT press, 1999.

[16] Dagum, L., and R. Menon. "OpenMP: an industry standard API for shared-memory programming." IEEE computational science and engineering 5, no. 1 (1998): 46-55.

[17] Nvidia, C.U.D.A. "Compute unified device architecture programming guide." (2007).

[18] Zaharia, M. et al. "Apache spark: a unified engine for big data processing." Comm. of the ACM 59, no. 11 (2016): 56-65.

[19] Dean, J., and S. Ghemawat. "MapReduce: simplified data processing on large clusters." Comm.of the ACM 51, no. 1 (2008): 107-113.

[20] Abadi, M. et al. "Tensorflow: a system for large-scale machine learning." In OSDI, vol. 16, pp. 265-283. 2016.

[21] Al-Rfou, R. et al. "Theano: A Python framework for fast computation of mathematical expressions." arXiv preprint (2016).

[22] Jia, Y. et al. "Caffe: Convolutional architecture for fast feature embedding." In Proc. of the 22nd ACM int. conference on Multimedia, pp. 675-678. ACM, 2014.

[23] Paszke, A. et al. "Automatic differentiation in pytorch." (2017).

[24] Meng, Xiangrui, Joseph Bradley, Burak Yavuz, Evan Sparks, Shivaram Venkataraman, Davies Liu, Jeremy Freeman et al. "Mllib: Machine learning in apache spark." The Journal of Machine Learning Research 17, no. 1 (2016): 1235-1241.

[25] The DeepHealth project, http://www.deephealth-project.eu

[26] Wickham, Hadley. "Tidy data." *Journal of Statistical Software* 59.10 (2014): 1-23.

[27] https://spack.io/

[28] https://easybuilders.github.io/easybuild/

[29] https://www.linux-kvm.org/

[30] https://www.docker.com/

[31] https://singularity.lbl.gov/

[32] https://aws.amazon.com/lambda/

[33] https://cloud.google.com/functions/

[34] https://openwhisk.apache.org/

[35] Howison, J. and Bullard, J., 2016. Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. *Journal of the Association for Information Science and Technology*, *67*(9), pp.2137-2155.

[36] Teredesai, A., et al., 2019. Machine learning model repository. U.S. Patent Application 10/275,710.

[37] R. H. Dennard, et al. Design of ion-implanted MOSFET's with very small physical dimensions. Solid-State Circuits, IEEE Journal of , 9(5):256–268, 1974.

[38] The Green500 list, https://www.top500.org/green500/

[39] Microsoft Catapult project, https://www.microsoft.com/en-us/research/project/project-catapult/

[40] Google's Tensor Flow Processing Unit, https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu Last access 2018-03-26

[41] https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html

[42] Piascik, R., et al., *Technology Area 12: Materials, Structures, Mechanical Systems, and Manufacturing Road Map*. 2010, NASA Office of Chief Technologist.

[43] https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120008178.pdf

[44] https://new.siemens.com/global/en/company/stories/industry/the-digital-twin.html

[45]A. Main, G. Scovazzi,The shifted boundary method for embedded domain computations. Part I: Poisson and Stokes problems,Journal of Computational Physics, 372,2018,972-995,https://doi.org/10.1016/j.jcp.2017.10.026.

[46]Song, T., et al.., 2018. The shifted boundary method for hyperbolic systems: Embedded domain computations of linear waves and shallow water flows. *Journal of Computational Physics*, *369*, pp.45-79.

[47] Nouveau, L., et al., 2019. High-order gradients with the shifted boundary method: An embedded enriched mixed formulation for elliptic PDEs. *Journal of Computational Physics*, *398*, p.108898.

[48] R. Zorrilla, et al.,A modified Finite Element formulation for the imposition of the slip boundary condition over embedded volumeless geometries, Computer Methods in Applied Mechanics and Engineering, 353,2019, 123-157

[49] "Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems", JCP, 262, 358–378

[50] *"Certified Reduced Basis VMS-Smagorinsky model for natural convection flow in a cavity with variable height"*, F. Ballarin, et al.

[51] *"Data-Driven POD-Galerkin Reduced Order Model For Turbulent Flows"*S. Hijazi, et al.

[52] "*Certified reduced basis methods and a posteriori error bounds for parametrized PDEs*", J. Hesthaven, et al, Springer 2015, and "*Model Reduction of Parametrized Systems*", P. Benner, et al., Springer, 2017

[53] "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions" N. Halko, et al. , 2009.

[54] "Fixed-precision randomized low-rank approximation methods for nonlinear model order reduction of large systems" C. Bach and F. Duddeck, 2019.

[55] "Model Order Reduction for Nonlinear Structural Dynamics Simulation-free Approaches", J.B. Rutzmoser, 2017.

[56] "*Nonlinear dimensionality reduction", 2007, J.A. Lee and M. Verleysen, Springer Verlag.*

[57] "Nonlinear manifold learning for model reduction in finite elastodynamics", D. Millan and M. Arroyo, 2013.

[58] "*Some themes in nonlinear model reduction with applications to structural dynamics"*, Jain, S. Ph.D. thesis, 2019. ETH Z.

[59] *"Efficient Geometrical parametrization for finite-volume based reduced order methods*", G.Stabile, G. Rozza, arXiv:1901.06373

[60] "*Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting". Farhat, C., et al., 2014. International Journal for Numerical Methods in Engineering 98 (9), 625–662.*
for computational efficiency

[61] "*Dimensional hyper-reduction of nonlinear finite element models via empirical cubature"*. 2017, Hernández, J.A., et al., Computer Methods in Applied Mechanics and Engineering 313, 687–722.

[62] E. N. Karatzas, et al.,A reduced basis approach for PDEs on parametrized geometries based on the shifted boundary finite element method and application to a Stokes flow, Computer Methods in App. Mechanics and Engineering, 347, 2019, 568-587

[63] Karatzas E.N., et al.. (2020) A Reduced Order Approach for the Embedded Shifted Boundary FEM and a Heat Exchange System on Parametrized Geometries. In: Fehr J., Haasdonk B. (eds) IUTAM Symposium on Model Order Reduction of Coupled Systems, Stuttgart, Germany, May 22–25, 2018. IUTAM Bookseries, vol 36. Springer, Cham

[64] H. Oliver et al., "Workflow Automation for Cycling Systems: The Cylc Workflow Engine", Computing in Science & Engineering Vol 21, Issue 4, July/Aug 2019. DOI: 10.1109/MCSE.2019.2906593.

[65] M Plóciennik, et al. “Two-level Dynamic Workflow Orchestration in the INDIGO DataCloud for Large-scale, Climate Change Data Analytics Experiments”, ICCS 2016, pp. 722-733.
[66] S. Fiore et al., “Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system”, BigData 2016, pp. 2911-2918.
[67] S. Fiore, et al., “An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management”, IEEE Access 7: 117652-117677 (2019)
[68] Wilkinson, M., et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, (2016)
[69] J. Macías, et al., "Performance benchmarking of Tsunami-HySEA model for NTHMP’s inundation mapping activities", Pure Appl. Geophys., vol. 174(8), pp 3147-3183, 2017.
[70] J. Selva, et al., Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA), Geophysical Journal International, Volume 205, Issue 3, 1 June 2016, Pages 1780–1803
[71] https://www.ndbc.noaa.gov/dart.shtml
[72] F. Løvholt, et al. (2019) Urgent tsunami computing, In 2019 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC)
[73] https://usgs.github.io/shakemap/sm4_index.html

Section 4: Members of the Consortium

4.1. Participants Profile

4.1.1 Barcelona Supercomputing Center – Centro Nacional de Supercomputación.

Short name	BSC
Type	Research Centre
Website	www.bsc.es
Country	Spain

Description of the organization

BSC is at the service of the international scientific community and of industry that requires HPC resources. Our multidisciplinary research team and our computational facilities –including MareNostrum– make BSC an international centre of excellence in e-Science. Since its establishment in 2005, BSC has developed an active role in fostering HPC in Spain and Europe as an essential tool for international competitiveness in science and engineering. The centre manages the Red Española de Supercomputación (RES), and is a hosting member of the Partnership for Advanced Computing in Europe (PRACE) initiative. We actively participate in the main European HPC initiatives, in close cooperation with other European supercomputing centres. With a total staff of more than 600 R&D experts and professionals, BSC has been successful in attracting talent, and our research focuses on four fields: Computer Sciences, Life Sciences, Earth Sciences and Computer Applications in Science and Engineering.
Most of BSC's research lines are developed within the framework of European Union research funding programmes, and the centre also does basic and applied research in collaboration with leading companies such as IBM, Microsoft, Intel, Nvidia, Repsol and Iberdrola. The centre has been extremely active in the EC Framework Programmes and has participated in seventy-nine projects funded by it. BSC is a founding member of HiPEAC, the ETP4HPC and participates in the most relevant international roadmapping and discussion forums and has strong links to Latin America. Education and Training is a priority for the centre and many of BSCs researchers are also university lecturers. BSC offers courses as a PRACE Advanced Training Centre, and through the Spanish national supercomputing network among others. The quality of our investigation has been recognized by the Spanish government with the Severo Ochoa Excellence Centre grant for cutting edge Spanish science.

The BSC-CNS Computer Sciences Department focuses on building upon currently available hardware and software technologies and adapting these technologies to make efficient use of supercomputing infrastructures. The department proposes novel architectures for processors and memory hierarchy and develops programming models and innovative implementation approaches for these models as well as tools for performance analysis and prediction.

The proposal is led by the Workflows and Distributed Computing group, which does research around the PyCOMPSs/COMPSs programming model, which makes it easier to develop workflows and parallel applications in distributed computing platforms. PyCOMPSs/COMPSs is used in production in MareNostrum 4 to execute scientific workflows. The group also does research on new storage systems, and deploys its ideas on the dataClay and Hecuba/Qbeast libraries. The group is very active in EC funded projects, with key contributions in projects such as ExaQUte, NEXTGenIO or the CoE BioExcel to mention a few related to the proposal.

BSC-CASE is a department focusing on developing software for the scientific and industrial communities. It covers many fields of expertise such as computational fluid dynamics, nuclear fusion or geophysics. Currently it hosts about 90 researchers among IT experts, physicists, mathematicians and other experts. The department has succeeded at establishing several industry R&D contracts (e.g. Repsol, Iberdrola, Rolls Royce, Siemens, SEAT). Currently it is broadening its areas of interest by including groups related to data analytics, quantum computing and risk analysis. It also hosts the coordinator of the ChEESE Center of Excellence for Solid Earth.

The Department of Earth Sciences of the Barcelona Supercomputing Centre-Centro Nacional de Supercomputación (BSC-CNS), BSC-ES henceforth (bsc.es/earth-sciences) is one of the most active groups in air quality and atmospheric composition modelling, climate prediction and climate services in Europe. The department is currently composed of about 100 people, including technical and support staff, structured in four distinct but interacting research groups: the Earth System Services, Atmospheric Composition, Climate Prediction and Computational Earth Sciences. The common vision and major scientific interests to all four groups are gathered under a departmental strategic plan, which is updated every second year.

The BSC-ES mission is performing research on and developing methods for environmental forecasting, with a particular focus on the atmosphere-ocean-biosphere system. This includes managing and transferring technology to support the main societal challenges through the use of models and data applications in high-performance computing (HPC) and data oriented infrastructures. It also includes the dissemination of real-time air quality and climate information based on its research expertise in collaboration with both the Spanish authorities and the World Meteorological Organisation (WMO).

Role of Partner in the Project

Tasks in the project: BSC coordinates the project, and leads WP1, WP6, WP8 and WP9.

Previous relevant expertise: The Workflows and Distributed Computing group (BSC-WDC) will contribute to the project with its expertise in workflow development with the PyCOMPSs environment, as one of the key tools in the project to develop applications that combine HPC, HPDA and AI. The expertise is not only at the programming model level, but also with expertise in complex runtime systems that enable the efficient execution in heterogeneous and distributed computing platforms. Another key component contribute by BSC-WDC is the machine learning library *dislib*, which offers an interface similar to scikit-learn, but parallelized internally with PyCOMPSs. BSC-WDC also contributes as expert in persistent storage systems, with the dataClay and Hecuba/Qbeast libraries. These components will be key to develop the eFlows4HPC workflows stack delivered by WP1, WP2 and WP3, that will be used in the development of the pillars' use-cases in WP4, WP5 and WP6.

BSC-CASE will coordinate Pillar III and contribute in the development of the UCIS4EQ workflow of urgent earthquake computing. The group has spearheaded the use of urgent HPC related to earthquake simulations for disaster resiliency. BSC-CASE has a long experience in production-grade software packages and in geoscience-related HPC modelling.

The BSC Earth-Sciences department, through the Computational Earth Sciences group, will be part of the Pillar II developing new workflows to run complex climate simulations. The group will bring expertise on the climate science itself but also in the computational aspects like workflow design and implementation, high-performance computing applied to earth system modelling, efficient analysis and handling of big I/O workloads and AI applications to earth science problems.

Expected benefits: BSC-WDC expects to extend its current workflow technologies based in PyCOMPSs into the much ambitious and complex eFlows4HPC software stack, which will include the storage libraries Hebuba/Qbeast and dataClay, and the machine learning library dislib. Together with the development and provisioning of the HPCWaaS methodology, BSC-WDC aims at positioning themselves as European leaders in workflow environments.

BSC-ES expects to develop improved workflows to apply in current climate simulations performed by the Climate Prediction group at BSC-ES. In addition, the data analysis and knowledge extraction developed in Pillar II, will be used by the Earth System Services team to produce better and faster climate services to deliver to end users.

BSC-CASE aims at bringing to an operational level its earthquake simulation workflow (UCIS4EQ) as well as making a strong case at operational and managerial levels to adopt urgent HPC policies in the European HPC ecosystem. We aim at obtaining the best possible damage estimates after the occurrence of earthquakes in areas of interest.

Qualification of key personnel

Dr Rosa M. Badia (female) holds a PhD on Computer Science (1994) from the Technical University of

Catalonia (UPC). She is the manager of the Workflows and Distributed Computing group at the BSC. Her current research interests include programming models for distributed computing platforms and its integration with novel storage technologies. She has participated in a large number of European projects (more than 30). The group lead by Dr Badia has been developing StarSs programming model for more than 10 years, with high success in adoption by application developers. Currently, the group focuses its efforts in PyCOMPSs/COMPSs, an instance of the programming model for distributed computing, with the goal of building workflows that combine computational, analytics and machine learning aspects. The initial target platforms for PyCOMPSs/COMPSs were traditional large clusters and cloud computing platforms. More recently this has been extended to include mobile devices and edge devices, with the objective of orchestrating the execution of workflows in fog-to-fog and fog-to-cloud scenarios. Rosa Badia is also involved in international strategic initiatives as the BDEC2 that focuses on platforms that include edge devices, HPC and cloud, for applications combining computation with analytics and artificial intelligence. Dr Badia has been IP of project SIENA, and of project EU-Brazil OpenBIO. She is also a member of the HiPEAC2 NoE. She is currently participating in EU funded projects: BioExcel2 CoE, CLASS, ExaQUte, mf2C, LANDSUPPORT and the EXPERTISE ETN, and in a contract with Fujitsu. Rosa M Badia received last year the Euro-Par Achievement Award 2019 and the DonaTIC Academia 2019.

Dr. Jorge Ejarque (male) holds PhD on Computer Science (2015) from the UPC. From 2005 to 2008 he worked as research support engineer at the UPC, and joined BSC at the end of 2008. He has contributed in the design and development of different tools and programming models for complex distributed computing platforms. He has published over 30 research papers in conferences and Journals and he has been involved in several National and European R&D projects (FP6, FP7 and H2020) as work package leader. He has been a member of several international conferences program committees, reviewer of journal articles and he was a member of the Spanish National Grid Initiative panel. His current research interests are focused on introducing energy efficiency in parallel programming models for heterogeneous parallel distributed computing environments and the interoperability between distributed computing platforms.

Dr. Anna Queralt (female) is a senior researcher at the Barcelona Supercomputing Center (BSC) since 2012. She holds a PhD in Computer Science (2009) from the Universitat Politècnica de Catalunya (UPC). She was a faculty member at the UPC from 2003 to 2012, coordinator and lecturer in the Technology for Big Data course at ESADE Business School from 2017 to 2018, and is lecturing in master and postgraduate courses on Big Data Management at UPC School for Professional and Executive Development since 2015. She was a member of the Steering Committee of the Standard Performance Evaluation Corporation (SPEC) Research Group, and a member of the SPEC Big Data Working Group. Her research interests are related to data sharing, storage systems, data management in edge to cloud, and the integration of data in the programming model to facilitate the development of data-intensive applications. She has served as a chair, organizer, and program committee member of several international conferences and workshops, and has participated in national and European projects related to these areas such as EUDAT, BigStorage, EXPERTISE, mF2C, CLASS or ELASTIC. The research line she leads has been developing the dataClay distributed object storage platform for the last 6 years, being evolved and exploited in the projects mentioned before.

Dr. Yolanda Becerra (female) holds a PhD in Computer Science since 2006 from the Technical University of Catalonia (UPC). She is an associate professor at the Computer Architecture Department of the UPC and an associate researcher in the Barcelona Supercomputing Center (BSC). Currently, she is part of the Workflows and Distributed Computing research group of the BSC, and leads the Data-driven scientific computing research line. The main goal of this line is to create a set of tools that boosts the performance of the storage system for big data scientific applications and that offers an easy and platform-independent storage interface.The tools developed include Hecuba, which implements strategies to offer a simple and efficient interface, and Qbeast, which implements an efficient and scalable multidimensional indexing and sampling mechanism.

Dr. Arnau Folch (male) Degree in Physics (University of Barcelona, 1994) and PhD in Applied Mathematics (Polytechnic University of Catalonia, 2000). Senior researcher at the CASE Department of BSC, where he leads the Environmental Simulations research group. Author of >80 scientific publications (h-index 29), he has participated in >30 national and European competitive research projects and in multiple contracts with private companies. Since 2011, he acts a the liaison officer between the International Union of Geophysics and Geodesy (IUGG) and the World Meteorological Organization

(WMO), and leads the Commission on tephra Hazard Modelling of the International Association of Volcanology and Chemistry of the Earth's Interior (IAVCEI). He has a wide expertise in ash dispersal forecasting and is one of the developers of the numerical model for ash transport FALL3D. He is also a permanent member of the Volcanic Ash Scientific Advisory Group (VASAG) of the IUGG-WMO. Having access to some of the most powerful supercomputing facilities in Europe, he has also developed a wide expertise in optimized computing strategies. He is the coordinator of the ChEESE Center of Excellence for Solid Earth.

Dr. Josep de la Puente (male) is a Physics graduate from the Universitat de Barcelona. He holds a PhD in Natural Sciences from the Ludwig Maximillians Universität Munich, obtained at the Earth Sciences Departments. He is a specialist in Computational Geophysics, including seismic/EM modelling and inversion. Josep specifically works in developing and deploying geophysical imaging software for high-performance computing platforms. His expertise in high-performance applications and geophysical software development has led to participating in projects related to technology transfer totaling 11 million EUR in competitive funding and 8 million EUR in industrial funding. Technology transfer has been mostly to Repsol SA, but also to PEMEX, Bull SAS, GEOOP or TOTAL. He has also been awarded 5 academic projects, national and international, of which he is principal investigator. At present time he manages a multi-disciplinary team including engineers, mathematicians and physicists focusing on HPC applications for geosciences.

Msc. Kim Serradell (male), is Master on High-Performance Computing from the Facultat d'Informàtica de Barcelona (FIB-UPC). Currently, he is the manager of the Computational Earth Science (CES) group at the Earth Sciences department in the Barcelona Supercomputing Center (BSC). The CES group is a multidisciplinary team of 31 members with different IT profiles that interacts closely with all the other groups of the Earth Sciences Dept. The group has among its tasks providing help and guidance to the scientists with the technical issues related to their work and developing a framework for the most efficient use of HPC resources. In the last years, he has been in charge for the system administration of all the computational resources of the department and he was also responsible for supervising the operational runs of the NMMB/BSC-Dust model and CALIOPE Air Quality System in the HPC infrastructures of the BSC. In that sense, he was also involved in the analysis of the models to improve their performance and developed strong skills of compilation and scripting. Furthermore, he's focused on deploying different earth system models (dust transport, climate or weather forecast) required by the department in a wide range of HPC architectures. He succeeds porting different these models in next HPC architectures like Montblanc cluster (ARM Based). He applied with success these skills in projects like IS-ENES2, IS-ENES3, ESiWACE, ESIWACE2, ESCAPE-2, SDS-WAS or BDFC.

Dr. Vanessa Fernandez (female) has more than 10 years' experience in R&D Project Management, with experience on the coordination and economic management of national and international projects primarily from European Commission Funding Programmes (H2020, FP7). She was project coordinator of the CORTICONIC project and participated in different actions and SPs at the FET FLAGSHIPS, HUMAN BRAIN PROJECT (SGA1 and SGA2) and NEUROGRAPHENE, GRAPHENE-CORE 1 and GRAPHENE CORE 2. She joined the BSC-CNS to provide support to several Computer Sciences Research Groups related to HPC (Heterogeneous architecture, programming models and performance tools). She is currently working as Project Manager for coordinating CLASS and ELASTIC projects. She holds a PhD in Neuroscience and Molecular Biology and a BA in Biology and Biotechnology.

Publications/achievements relevant to the call

- Tejedor, Enric, Yolanda Becerra, Guillem Alomar, Anna Queralt, Rosa M. Badia, Jordi Torres, Toni Cortes, and Jesús Labarta. "Pycompss: Parallel computational workflows in python." The International Journal of High Performance Computing Applications 31, no. 1 (2017): 66-82.
- Cid-Fuentes, Javier Álvarez, Pol Álvarez, Ramon Amela, Kuninori Ishii, Rafael K. Morizawa, and Rosa M. Badia. "Efficient development of high performance data analytics in Python." Future Generation Computer Systems (2019).
- Conejero, Javier, Cristian Ramon-Cortes, Kim Serradell, and Rosa M. Badia. "Boosting Atmospheric Dust Forecast with PyCOMPSs." In 2018 IEEE 14th International Conference on e-Science (e-Science), pp. 464-474. IEEE, 2018. Harvard

- Bellprat, O., Guemas, V., Doblas-Reyes, F., Donat, M. G., "Towards reliable extreme weather and climate event attribution". Nature Communications, 10, 1732 (2019) doi:10.1038/s41467-019-09729-2.

- Boehm, C., Hanzich, M., de la Puente, J. and Fichtner, A., "Wavefield compression for adjoint methods in full-waveform inversion". Geophysics 81 (6), R385-R397, 29 (2016) doi:10.1190/geo2015-0653.1.

Relevant previous projects/activities

ChEESE: Coordinated by BSC, ChEESE will harness European institutions in charge of operational monitoring networks, tier-0 supercomputing centers, academia, hardware developers and third-parties from SMEs, Industry and public-governance. The scientific ambition is to prepare 10 flagship codes to address Exascale Computing Challenging (ECC) problems on computational seismology, magnetohydrodynamics, physical volcanology, tsunamis, and data analysis and predictive techniques for earthquake and volcano monitoring. ChEESE will promote and facilitate the integration of HPC services to widen the access to codes to the Solid Earth users community. Finally, ChEESE aims at acting as a hub to foster HPC across the Solid Earth Community and related stakeholders and to provide specialized training on services and capacity building measures.

ESiWACE: The Centre of Excellence in Simulation of Weather and Climate in Europe (ESiWACE) enables global storm- and eddy resolving weather and climate simulations on the upcoming (pre-)Exascale supercomputers. By explicitly resolving storm dynamics, it will vastly improve weather and climate simulations. These results will inform decisions on emission reductions as well as adaptation strategies for housing, cities, farming, coastal defenses and other parts of society.

ExaQUte: The ExaQUte project aims at constructing a framework to enable Uncertainty Quantification and Optimization under Uncertainties in complex engineering problems, using computational simulations on Exascale systems. In the framework of ExaQUte, BSC is extending PyCOMPSs to support Multi-Level Monte-Carlo workflows to solve engineering problems.

BioExcel2: BioExcel is the European Centre of Excellence for provisioning support to academic and industrial researchers in the use of high-performance computing (HPC) and high-throughput computing (HTC) in biomolecular research. The BSC group is providing PyCOMPSs in BioExcel for the development of workflows that combine molecular dynamics and data analytics.

USER: EuroLab4HPC Business Proposal Prototyping validates the feasibility of urgent earthquake simulations not only for civil protection but for industrial applications, in particular in the business of reinsurance and earthquake engineering.

Relevant infrastructures and networks

BSC hosts the MareNostrum 4 supercomputer, a general-purpose block with 3,456 homogeneous nodes, total of 165,888 cores. Part of MareNostrum 4 is also and an emerging technology cluster: an IBM Power9 processors with NVIDIA Volta GPUs. These infrastructures can be used for BSC internal research activities, and BSC will support the partners on granting access through peer-reviewed processes.

4.1.2. Centre Internacional de Mètodes Numèrics a l'Enginyeria

Short name	CIMNE
Type	Research Centre
Website	http://www.cimne.com
Country	Spain

Description of the organization

The International Center for Numerical Methods in Engineering (CIMNE) was founded in 1987 at the heart of the Technical University of Catalonia (UPC) under the auspices of UNESCO. CIMNE is an autonomous research and development centre currently employing around 170 scientists from 25 different countries worldwide.

CIMNE has been recently awarded with the "Severo Ochoa Centre of Excellence". With this call, the Ministry of Science, Innovation and Universities aims to promote high-impact research carried out in the R&D centres of Spain. The accredited centres stand out both for the international notoriety of the scientific contributions they make, and for their innovative capacity and their intense relationship with the business sector. They are also world reference centres capable of attracting international talent. This Excellence Award recognized the international leadership of CIMNE and positions our centre in the lists of top research centres in Spain for its outstanding role in the development of numerical methods and its application to many different fields in science and technology.

The research activities of CIMNE aim of promoting and fostering advances in the development and application of numerical methods and computational techniques for the solution of engineering problems in an international context. CIMNE organizes a wide range of activities aimed at teaching and spreading of knowledge in the field of computational engineering, such as courses, seminars, conferences and publications. Additionally, CIMNE carries out various research and development activities and has participated in a large number of technology transfer projects in cooperation with many enterprises and organizations from different countries. CIMNE has participated in around 150 projects of EC programmes and has acted as coordinator in 40 of these projects.

The research activities of CIMNE cover the development of innovative constitutive models for composite materials and structures, new numerical methods for non-linear analysis and safety studies of structures, shape optimization in structural and fluid dynamic problems, computational fluid dynamics studies for both external and internal flow problems and numerical simulation of material deformation and forming processes for the manufacturing industry, mesh generation and visualization interfaces, casting and thermal process, stochastic optimization as well as program parallelization and distributed (grid and cloud) computing techniques. Such know-how has been applied successfully to the development of Reduced Order Models to be used in multiple fields of engineering and of material modelling.

Role of Partner in the Project

CIMNE will act as coordinating partner of the "Manufacturing Pillar" (WP4) of the proposal. Within the pillar, will provide a leading expertise in the development of high-performance solvers for the solution of problems in manufacturing, as well as lead the development of the Kratos open source platform which will be used for the developments. To this end, the project team includes in-house experts in the field of HPC and of Reduced Order modelling. The team will be complemented by a visiting scientist, Prof. Guglielmo Scovazzi who will work at CIMNE during the summers so to provide expertise in embedded methods and in numerical analysis.

Within the pillar, CIMNE team will work towards the definition of a Digital Twin to be used both in the design phase and as a life companion of manufactured products. The main goal will be the development of techniques to allow the use of large-scale HPC hardware in the training of Reduce Order Models to be later deployed on edge-class hardware.

One of the main research interests will be the integration of the solution capabilities available in the group with BSCs workflow management systems. Such integration will allow to take advantage of the project

developments in the field of Machine Learning and Artificial Intelligence.

Qualification of key personnel

Dr. Riccardo Rossi (male) is a Civil Engineer from the University of Padova (2001). He obtained his PhD at the University of Bologna in 2005. He is currently associate professor (Prof. Agregat) at the UPC, and Full Research Professor at CIMNE, where he works as affiliates scientist since 2005. During his career he collaborated actively both in the field of Computational Fluid Dynamics (CFD) and in Computational Structural Mechanics (CSD). He is one of the main developers of the general-purpose code "Kratos" (of which he holds copyright) where he is one of the responsibles of the CFD part and of the "StructuralMechanics" solver. To date he has published more than 50 indexed JCR papers, 6 book chapters as well as 6 Monographes. He has directed 9 completed PhD theses. To date he has an h-factor of 19 (Scopus) or 23 (Google Scholar). He has a long expertise in the field of HPC, where he coordinated the ExaQUte project and participated as responsible of CFD in the Numexas project. He also worked very actively in the field of solution processes for manufacturing, and was the main developer of the solver employed as the basis of the "Click2Cast" commercial solver. In the last three years he had ongoing industrial projects in the field of simulation for manufacturing for a total budget of more than 200k€/year. Dr. Rossi is currently one of the guarantors of CIMNE's Severo Ochoa excellence grant, and leads the "Kratos Multiphysics" research group within the institution.

Dr. Joaquín A. Hernández (male) is a Mechanical Engineer (2003) from the Technical University of Cartagena, Spain. He obtained is PhD at the Technical University of Catalonia (UPC) in 2009. He is currently Associate professor at UPC, in the School of Industrial and Aerospace Engineering of Terrassa, and Full Research Professor at CIMNE. His research revolves around the field of Computational Mechanics. His first contributions to this field were on numerical modeling of flow and compaction of granular/powder materials. Since 2012, his research is focused on the synergistic combination of Multiscale Techniques with Reduced-Order Modeling, with special emphasis on the issue of fast integration of governing equations (or ``hyperreduction"). Dr. Joaquín A. Hernández has published 16 JCR papers, 1 book, 2 book chapters and 5 monographs. To date, he has h-factor of 11 (Scopus).

Dr. Guglielmo Scovazzi (male) is Professor of Civil & Environmental Engineering and Mechanical Engineering & Materials Science at Duke University (USA). His interests are in the general area of computational mechanics, and specifically in CFD, computational solid mechanics, fluid/structure interaction, computational geomechanics, flow through porous media. He earned B.S./M.S. Degrees in Aerospace Engineering at Politecnico di Torino in 1998. He received a M.S. in 2001 and a Ph.D. in 2004, both from the Mechanical Engineering Department at Stanford University. Between 2004 and 2012, he held a position as Senior Member of the Technical Staff at Sandia National Laboratories, Albuquerque (New Mexico). Guglielmo Scovazzi is a recipient of the 2014 Early Career Award from the Office of Science of the US Department of Energy (ASCR program), and the 2017 Presidential Early Career Award for Scientists and Engineers (PECASE), the highest honor bestowed by the US Executive Office (The White House) to research scientists in the Early stages of the academic careers. In February 2018, he was named Kavli Fellow by the National Academy of Sciences and the Kavli Foundation. The Kavli Fellowship acknowledges contributions of exceptional U.S. scientists under the age of 45. He is an associate editor of the Journal of Computational Physics, and in the editorial board of the International Journal on Numerical Methods for Fluids. He is a member of the executive board of USACM and a member of SIAM and ASME.

Mr. Ruben Zorrilla (male) is a civil engineer and Master in Numerical Methods from UPC. He is currently in the second year of his PhD focused on the development of a "Virtual Wind Tunnel". His current work focuses on the simulation of FSI problems employing an embedded CFD technology. He has expertise both in OpenMP and MPI programming.

Dr. Cecilia Soriano (female) holds a PhD in Physics from the Technical University of Catalonia (UPC), with a specialization in Environmental Engineering and meteorological and dispersion air pollution modeling. She is a part term lecturer of Mechanics at the School of Civil Engineering of the UPC and she is the Research Manager at CIMNE. She has long experience in scientific project management and follow-up.

Mr. Carlos Roig (male) is a computer scientist, and master in Research and Innovation from UPC. He has a multi year experience in programming software for HPC, employing a variety of programming models, including MPI, OpenMP and CUDA. He is also expert in the profiling of HPC applications using EXTRAE/PARAVER and VTUNE. He has been one of the primary developers in the Symphony project.

Mr. Raul Bravo (male) is a Mechanical Engineer from the National Institute of Technology of Mexico (2016) and a Master of Science in Computational Mechanics from Swansea University-Technical University of Catalonia (2019). He is in the first year of his PhD at CIMNE, where he is working on the topic of Reduced Order Modelling with application to Digital Twins. He is one of the main developers of the ROM Application of Kratos.

Publications/achievements relevant to the call

- Hernández, J.A., Caicedo, M.A., Ferrer, A., "Dimensional hyper-reduction of nonlinear finite element models via empirical cubature". 2017, Computer Methods in Applied Mechanics and Engineering 313, 687–722
- R. Zorrilla, A. Larese, R. Rossi, "A modified Finite Element formulation for the imposition of the slip boundary condition over embedded volumeless geometries", Computer Methods in Applied Mechanics and Engineering,353,2019, 123-157
- S. Zaghi, X. Martinez, R. Rossi, M. Petracca, "Adaptive and off-line techniques for non-linear multiscale analysis", Composite Structures, 206,2018, 215-233
- E Soudah, R Rossi, S Idelsohn, E Oñate, "A reduced-order model based on the coupled 1D-3D finite element simulations for an efficient analysis of hemodynamics problems", Computational Mechanics 54 (4), 1013-1022
- F Salazar, R Morán, R Rossi, E Oñate, "Analysis of the discharge capacity of radial-gated spillways using CFD and ANN–Oliana Dam case study", Journal of Hydraulic Research 51 (3), 244-252

Relevant previous projects/activities

NUMEXAS: Numerical Methods and tools for key exascale computing challenges in engineering and applied sciences.FP7-ICT, 2013-2016.

VELaSSCo: Visualization for Extremely Large-Scale Scientific Computing, FP7-ICT, 2014-2016.

SimPhoNy: Simulation framework for multi-scale phenomena in micro- and nanosystems. FP7-NMP, 2014-2016

CaXMan: Computer Aided Technologies for Additive Manufacturing. H2020, 2015-2018.

ExaQUte: EXAscale Quantification of Uncertainties for Technology and Science Simulation. H2020, 2018-2021.

Relevant infrastructures and networks

CIMNE has its own in-house cluster for parallel computing, Acuario, and also has access to the computing facilities of several supercomputer centers in US, Japan, Europe and Spain (including BSC, CSUC, CESGA, etc.).

4.1.3. Forschungszentrum Jülich GmbH

Short name	FZJ	
Type	Research Centre	
Website	www.fz-juelich.de	
Country	Germany	

Description of the organization

The Forschungszentrum Jülich (FZJ) – a member of the Helmholtz Association – is one of the largest research centres in Europe. It pursues cutting-edge interdisciplinary research addressing the challenges facing society in the fields of health, energy and the environment, and information technologies. The Jülich Supercomputing Centre (JUELICH, JSC) at Forschungszentrum Jülich operates one of the most powerful supercomputer infrastructures for scientific and technical applications in Europe and grants scientists in Germany and Europe access to these resources for their research. JSC operates storage systems for users of the HPC systems since the early 1980s, today JSC operates storage systems with more than 100 PB disk and 300 PB tape capacity. JSC's division "Federated Systems and Data" (FSD) has been involved in federated and grid computing as well as data management and analysis for many years. The division plays an active role in various national, EC-funded, and international R&D and infrastructure projects (to name a few: EOSC-hub, EUDAT CDI, HBP, PRACE, Helmholtz Data Federation and Helmholtz Analytics Framework).

Role of Partner in the Project

Tasks in the project: FZJ will lead Work Package 2 in the project, which is responsible for efficient deployment and optimization of the software stack defined in the WP1. The participant will also capitalize on its long-year software development experience by defining software architecture and integration plan in WP1. FZJ will also be active in implementing Pillars' use cases, especially with respect to defining data pipelines in Data Logistics Service and model repository. Lastly, JSC will be one of the resources providers to run the services and Pillars workflows.

Previous relevant expertise: Jülich Supercomputing Centre (JSC) was actively shaping the European landscape of scientific software and services in projects like EMI (Grid Computing), EUDAT, EUDAT2020 (data infrastructures), and remains a key player in implementing vision of European Open Science Cloud (EOSC-hub).

Expected benefits: FZJ is a world-leading actor in the area of modular supercomputing, such architectures can only be efficiently exploited with tools that support distributed workflows as developed in the eFlows4HPC project. The project will also provide the opportunity to widen the user-base of the FZJ internal tools like UNICORE, HeAT, HDF, Data Logistics Service, and thus improve their functionality range. On the other hand, tools and software suits provided by project partners, e.g., pyCOMPS, DataClay fit well in the JUELICH portfolio and cover crucial aspects of scientific workflows implemented in JUELICH.

Qualification of key personnel

Daniel Mallmann (male) is head of the division "Federated Systems and Data" (FSD) in the Jülich Supercomputing Centre (JSC) of Forschungszentrum Jülich since 2010. He is a graduate electrical engineer by training, received his diploma from the RWTH Aachen. Daniel is involved in national, European and international projects since 2000. Currently, he is a member of the EUDAT CDI Executive Board and of the EOSC-hub Project Management Board. The division is the major development partner of the federation software suite UNICORE and contributes to infrastructures including EUDAT, PRACE, and the Human Brain Project. His major research interests are in distributed systems, especially infrastructure federations, data management, virtualization, and security.

Dr. Jedrzej Rybicki (male) received the doctoral degree in computer science in 2011 from Heinrich Heine University in Düsseldorf. Subsequently, he moved to Jülich Supercomputer Center to join the group of Federated Systems and Data where he works on data-intensive, distributed processing in Projects EMI,

EUDAT, EUDAT2020. His major research interests are software architectures for data-intensive applications.

Dr. Bernd Schuller (male) holds a doctoral degree in Physics from the Technical University, Aachen (Germany). Since 2003 he has been working at Juelich Supercomputing Centre (Germany) and was involved in many research projects as a scientist and software architect, most recently the Human Brain Project (HBP). He is one of the main developers of the UNICORE middleware (https://www.unicore.eu). His main research interests are distributed and federated systems, workflow systems, data management and security.

Björn Hagemeier (male) received his diploma in computer science from the University of Paderborn, Germany in 2006. Since then, he has been working at Jülich Supercomputing Centre, where he was involved in numerous projects. Björn currently heads the group of Data Services and Infrastructure within the division Federated Systems and Data at JSC. As part of his responsibilities, his personal work involves the operation of the Helmholtz Data Federation cloud resources at JSC. He is project manager of the Helmholtz Analytics Framework project funded within the Helmholtz Initiative and Networking Funds. His main research interests are distributed and federated systems, IaaS cloud, data management, data services, and machine learning.

Publications/achievements relevant to the call

- Rybicki, J. "Application of Event Sourcing in Research Data Management", The Fourth International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2018), http://hdl.handle.net/2128/18449.
- Schuller, B. ; Rybicki, J. ; Benedyczak, K. "High-Performance Computing on the Web: Extending UNICORE with RESTful Interfaces", The Sixth International Conference on Advances in Future Internet, IARIA XPS Press 35-38 (2014), ISBN: 978-1-61208-377-3
- Benedyczak, K. ; Schuller, B. ; Rybicki, J. et.al., "UNICORE 7 - Middleware Services for Distributed and Federated Computing", HPCS 2016, IEEE, 2016, 613-620. DOI 10.1109/HPCSim.2016.7568392
- K. Krajsek, C. Comito, M. Götz, B. Hagemeier, Ph. Knechtges, M. Siggel, "The Helmholtz Analytics Toolkit (HeAT) - A Scientific Big Data Library for HPC", Extreme Data – Demands, Technologies, and Services, Jülich, 18.-19. Sept. 2018, IAS Series Vol. 40, 2019, 57-60.
- B. Hagemeier, O. Bücker, A. Giesler, R. Saini, B. Schuller, "A Workflow for Polarized Light Imaging Using UNICORE Workflow Services", UNICORE Summit 2014, Leipzig, Germany, 24 Jun 2014, IAS Series Vol. 26, 2014, 1-14.

Relevant previous projects/activities

UNICORE (Uniform Interface to Computing Resources) provides tools and services for building federated systems, making high-performance computing and data resources accessible in a seamless and secure way for a wide variety of applications in intranets and the internet. https://www.unicore.eu/

HeAT. Developed within the Helmholtz Analytics Framework project (http://www.helmholtz-analytics.de), the HeAT toolkit combines the strengths of distributed parallelism predominant in the HPC domain with GPU supported parallelism predominant in the machine learning domain. HeAT is developed in the open and its source code is hosted on GitHub (https://github.com/helmholtz-analytics/heat).

PRACE. The Partnership for Advanced Computing in Europe (PRACE) is an international non-profit association with its seat in Brussels. The PRACE Research Infrastructure provides a persistent world-class high performance computing service for scientists and researchers from academia and industry in Europe. The computer systems and their operations accessible through PRACE are provided by 5 PRACE members (BSC representing Spain, CINECA representing Italy, ETH Zurich/CSCS representing Switzerland, GCS representing Germany and GENCI representing France). http://www.prace-ri.eu/

High Performance Analytics & Computing Platform of the Human Brain Project. The High Performance Analytics and Computing (HPAC) Platform provides neuroscientists with the high performance computing, storage and data processing capabilities they need to run simulation of sophisticated, detailed brain models and to analyse large, complex data sets. It also provides software tools

and frameworks for neural network simulations that run both on the user's computer and efficiently on HPC systems, for scientific visualisation, interactive and visual data analytics, performance analysis of parallel applications, data management, dynamic load balancing, parallel programming models, and numerical models for brain simulations. https://www.humanbrainproject.eu/en/massive-computing/
Relevant infrastructures and networks
JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre. Journal of large-scale research facilities, 5, A135. doi: 10.17815/jlsrf-5-171 HDF Cloud – Helmholtz Data Federation Cloud Resources at the Jülich Supercomputing Centre. Journal of large-scale research facilities, 5, A137. http://dx.doi.org/10.17815/jlsrf-5-173.

4.1.4. Universitat Politecnica De Valencia

Short name	UPV
Type	University
Website	www.upv.es
Country	Spain

Description of the organization

Universitat Politècnica de València (UPV, http://www.upv.es) is a public university with four campus sites, over 35,000 students, and 2,600 faculty members and research staff. It consists of 44 Departments, most of them in engineering areas. It is the top University in Spain regarding patent production.

Contributions to this project will come from the Parallel Architectures group (GAP, http://ww.gap.upv.es) of UPV. The group has a 26-year research expertise in different aspects of system architecture, especially on interconnection networks and architecture-aware optimization of algorithms. Currently, the group is focused in the modeling, optimization and application of machine learning techniques and algorithms in different architectures, including hardware accelerators and distributed memory platforms. Currently, the group is formed of twenty-nine researchers. Members of the team have recognized participation and contributions in top ranked conferences for Computer Architecture field and Parallel and Distributed Systems, being TPC members of ISCA, HPCA, SuperComputing conference, as well as Associate Editors of IEEE journals such as TC, CAL, and TPDS; ACM Trans. on Mathematical Software; and Elsevier Parallel Computing.

Role of Partner in the Project

Tasks in the project: UPV will lead WP3 (co-design aspects between applications, software stack and actual hardware). As WP leader, UPV will coordinate the actions by the participating partners toward the identification, development and integration of highly optimized kernels into the project applications. UPV will also support the application of machine learning techniques to the acceleration of the pillars.

Previous relevant expertise: Participant members of the GAP group have lead research projects focused on FPGA and on-chip interconnects, providing also virtualization and partitioning support in embedded multicore systems. The members have participated in several processor design initiatives and developed a multicore architecture (from the basic processor, to the networks and memory hierarchy) in Verilog and ready for being used in the FPGA-based prototyping solution. In the last years, the GAP has gained a strong expertise in the development of accelerator architectures and its emulation and prototyping in multi-FPGA boards. In addition, the group members have extensive expertise in the architecture-aware optimisation, including the parallelisation, of complex scientific applications. The group is codeveloping the EDDL (European Distributed DeepLearning) library in the framework of the EU H2020 DeepHealth project as well as participating with private companies in several contracts related with Machine Learning.

Expected benefits: The members of UPV have large experience in the architecture-aware optimization of algorithms, both in GPUs and FPGAs. The participation in this project will enlarge this experience to other relevant use cases, in particular, deep learning technologies, that improve the possibility of transferring this know-how to industrial sectors. The project benefits will complement the DeepHealth project and vice-versa, enabling cross fertilization. The large FPGA prototyping system owned by UPV will be fully exploited in the project, helping to spread among the research community the energy advantages of FPGAs.

Qualification of key personnel

José Flich Cardo (male) received his PhD in 2001 in Computer Engineering. He is Full Professor at UPV where he leads the research activities related to NoCs. He has served in different conference program committees (ISCA, PACT, NOCS, ICPP, IPDPS, HiPC, CAC, ICPADS, ISCC), as program chair (INA-OCMC, CAC) and track co-chair (EUROPAR, SC). José Flich Cardo has collaborated with different Institutions (Ferrara, Catania, Jonkoping, USC) and companies (AMD, Intel, Sun). His current research activities focus on reconfiguration, routing, coherency protocols and congestion management within NoCs. He has co-invented different routing strategies, reconfiguration and congestion control mechanisms, some of them with high recognition (RECN and LBDR for on-chip networks). He is a member of the HIPEAC

NoE. He has co-edited a book for the field of networks-on-chip. He was the Coordinator of the FP7 STREP NaNoC project and the FETHPC MANGO project. Currently he is involved in H2020 RECIPE project and H2020 DeepHealth projects.

Enrique S. Quintana-Ortí (male) is Professor of Computer Architecture at UPV. Enrique's research pursues the optimization of numerical algorithms and deep learning frameworks for general-purpose processors as well as hardware accelerators. During the past 22 years, he has co-authored 100+ papers in peer-reviewed scientific journals and 200+ in international conferences. He has also participated in international projects funded by European organizations (EU FP7 TEXT and Exa2GREEN, EU H2020 INTERTWinE and OPRECOMP), as well as in USA projects from DoE and NSF. Enrique's research on fault-tolerance has been recognized in by USA NASA with two awards, and his contributions to the acceleration of linear algebra algorithms received the 2008 NVIDIA Professor Partnership Award. Finally, he has served as member of the scientific committee of 80+ international conferences; he is area editor for Elsevier's Parallel Computing journal and he is Associate Editor for ACM Trans. on Mathematical Software.

Pedro López (GAP, male) received his Ph.D. degree in computer engineering in 1995. Since 2002, he is a Professor in the Department of Computer Engineering (DISCA) at UPV. His research interests mainly include processor microarchitecture, cache memory, high performance interconnection networks for multiprocessor systems and clusters and more recently neural networks and deep learning. He has published over 140 refereed conference and journal papers which received more than 3,300 citations and a h-index=32 (according to Google Scholar). He served in several conference program committees and especially in the editorial board of the Parallel Computing Journal for ten years. Currently he is involved in H2020 DeepHealth and H2020 Selene projects.

Carles Hernández Luz (male) is a senior Researcher at the Universitat Politècnica de València. Previously from 2012 to 2018 he was senior researcher at the CAOS group from Barcelona Supercomputing Center. In 2012 he worked as intern at the IP verification group at Intel Mobile Communications Munich. His area of expertise includes on-chip interconnects, processor design, real-time aware high-performance computing, and reliability. He is currently co-advising 5 PhD students. Dr. Hernandez is the project coordinator of H2020 SELENE project on High-performance and dependable computing. In 2015 he was granted with a Young Researcher Grant by the Spanish Ministry to conduct research on high-performance and reliable processor design. He was the PI of BSC activities in the FET-HPC RECIPE project on predictable heterogeneous Exascale computing and participates/has participated in H2020 EPI and DEEPHEALTH projects, NaNoC, parMERASA, PROXIMA IP7 and VeTeSS ARTEMIS projects.

Publications/achievements relevant to the call

- G. Flegar, F. Scheidegger, V. Novakovic, G. Mariani, A. E. Tomás, A. C. M. Malossi, E. S. Quintana-Ortí. "FloatX: A C++ library for customized floating-point arithmetic". ACM Trans. on Mathematical Software, 2019. To appear.

- S. Catalán, J. R. Herrero, E. S. Quintana, R. Rodríguez-Sánchez. "Energy balance between voltage-frequency scaling and resilience for linear algebra routines on low-power multicore architectures". Parallel Computing, Vol. 73, pp. 28-39, 2018.

- J. Flich et al: "Exploring manycore architectures for next-generation HPC systems through the MANGO approach". Microprocessors and Microsystems - Embedded Hardware Design 61: 154-170 (2018)

- G. Agosta et al: "Challenges in Deeply Heterogeneous High Performance Systems". DSD: 428-435 (2019)

Relevant previous projects/activities

EU H2020 732631 Project "**OPRECOMP**. Open Transprecision Computing". Coordinator: C. Bekas (IBM Zürich). January 2017-December 2020.

EU H2020 671602 Project "**INTERTWinE**. Programming Model Interoperability Towards Exascale". Coordinator: M. Bull (EPCC). October 2015 – September 2018.

EU H2020 **Deep-Learning** and HPC to Boost Biomedical Applications for Health, Coordinator: M. Caballero (EVERIS), January 2019 – December 2021. EU H2020 **RECIPE** (GA# 801137), Reliable power and Time-Constraints-Aware predictive management of heterogeneous exascale systems, Principal Investigator(s): W. Fornaciari (POLIMI) - May 2018 – April 2021. EU H2020 **SELENE** Self-monitored Dependable platform for Safety-Critical Systems, Coordinator: C. Hernández (UPV) – December 2019 – November 2022.
Relevant infrastructures and networks
MANGO Prototype made of 96 high-end FPGAs, interconnected pin to pin.

4.1.5. Bull SAS

Short name	ATOS	
Type	Industry	
Website	www.bull.com	
Country	France	

Description of the organization
ATOS is a global leader in digital transformation with 110,000 employees in 73 countries and annual revenue of over 12 billion. European number one in Cloud, Cybersecurity and High-Performance Computing, the Group provides end-to-end Orchestrated Hybrid Cloud, Big Data, Business Applications and Digital Workplace solutions through its Digital Transformation Factory, as well as transactional services through Worldline, the European leader in the payment industry. With its cutting-edge technologies and industry knowledge, Atos supports the digital transformation of its clients across all business sectors. The Group is the Worldwide Information Technology Partner for the Olympic & Paralympic Games and operates under the brands Atos, Atos Syntel, Unify and Worldline. Atos is listed on the CAC40 Paris stock index. With its deep technology expertise and industry knowledge, the Group works with clients across different business sectors: Defense, Financial Services, Health, Manufacturing, Media, Utilities, Public sector, Retail, Telecommunications, and Transportation. Atos Big Data Security (BDS), contributing to this project, is the business line for Atos' technology products and software (former BULL SAS). With a rich heritage of over 80 years of technological innovation, 2,000 patents and a 700 strong R&D team supported by the Atos Scientific Community, Atos BDS offers products and value-added software to assist clients in their digital transformation, specifically in the areas of Big Data and cyber security. Atos BDS is the European leader in High Performance Computing (HPC) and its products include the energy-efficient supercomputer that is based on a system patented by the company; Bullion, one of the most powerful x86 servers in the world developed to meet the challenges of Big Data.
Role of Partner in the Project
In the project Atos will contribute to the design and implementation of the software stack to support complex workflows that combine HPC, HPDA and AI (ML) workloads. Bull will focus on high level workflow management, leveraging its open source orchestration solution (YORC) to support the high-level application lifecycle management, i.e. deploying, configuring, starting, synchronizing the HPC/HPDA/AI components that constitute a workflow. Atos will adapt and extend the YORC orchestrator to address the project requirements that may include dynamic workflows, failover workflows, data aware workflows. In addition Atos will also contribute to artifacts re-use and management, in particular for the AI models management, Atos will leverage and enhance its AI Software stack dealing with ML models tuning, training, inference. These technical contributions will be mainly achieved in WP1, where Atos will lead the task T1.5. Atos will also lead the WP7 about exploitation. **Previous relevant expertise:** HPC, HPDA and AI (ML/DL); orchestration solution **Expected benefits:** This project will provide Atos the opportunity to enrich and enhance our offer Codex AI regarding the orchestration solution and AI models engineering. Moreover, Atos' know-how will be enriched by contributing to the project and by working with the partners of the project.
Qualification of key personnel
Dr. François Exertier (male): R&D manager at Atos BDS. He is technical leader of a AI & HPC R&D team dealing about HPDA (High Performance Data Analytics) and AI; the developed solutions are about Big Data and AI Application Design, Development, Lifecycle Management and Orchestration. He has experience in participating in many National and European collaborative projects (FUI, FP7, ITEA...). He has also experienced many open source projects, in the area of Java EE, SOA, Administration, Cloud Computing / PaaS (JOnAS, JASMINe projects). He got a PhD Thesis in Computer Science from University Joseph Fourier of Grenoble, France, in 1991. **Loïc Albertin** (male): Software Engineer at Atos BDS in the R&D department, working as developer for

solutions to design, develop, operate and orchestrate deployment of HPDA (High Performance Data Analytics) and more generally Big Data applications on top of HPC and Cloud infrastructures. He has a strong background in open-source development as a major contributor in several OW2 projects focused on Java EE, SOA and PaaS (JOnAS, EasyBeans, JASMINe). He is the technical leader of the open sourced orchestration solution named Yorc (https://github.com/ystia/yorc). He holds a Master Degree in Computer Science (2006) obtained from Joseph Fournier University, France.

Claire Chen (female), R&D project manager: Claire Chen has obtained her engineer degree in computer science from Telecom Paris in France. She worked as a software engineer, software project manager and product manager for data protection offers. Now in Atos/Bull R&D teams, she is a project manager for the innovation projects.

Etienne Walter (male) is senior project manager in Bull R&D Division, in charge of HPC projects: Etienne worked in Bull Mainframe division, as a software developer, as manager in charge of Change Control process and of a software development team. For 5 years he is involved in HPC and Big Data projects. He has been or is involved in several European projects, such as Servery (Celtic project), DataScale & ELCI (French Initiative), Fortissimo (FP7), Mont-Blanc2 (FP7) and is now coordinating Mont-Blanc 3 H2020 project.

Publications/achievements relevant to the call

HPC in Europe with the active leadership of ETP4HPC and contribution to the Strategic Research Agenda. Atos BDS has developed a complete open exascale-class supercomputer offer containing Hardware and Software technologies, BullSequana. With BullSequana, Atos delivers an innovative solution that matches the exascale technological challenges. The Atos R&D designed BullSequana around the following guidelines:

- Open and multi-technology
- Ultra-dense and scalable
- Ultra-energy efficient
- Easy administration

1. With the BullSequana X range of supercomputers, Atos confirms its strategic commitment to the development of innovative high-performance computing systems – the systems needed to meet the major challenges of the 21st century. Designed by Atos R&D in close cooperation with major customers, Bull Sequana X supercomputer leverages the latest technological advances, so as to guarantee maximum performance for a minimized operation cost. The race to exascale calls for technological breakthroughs.
2. BullSequana S series: To tackle enterprise IT challenges and enable businesses to take full advantage of Artificial Intelligence (AI), Atos brings to the market a new generation of x86 servers, BullSequana S, optimized for machine-learning, business critical computing applications and in-memory environments. BullSequana S reaches the highest level of quality of service, performance, availability and scalability to meet IT departments' existing and emerging demands.
3. BullSequana M series: Atos mainframe servers have always been able to evolve in order to integrate the latest technologies, especially to take into account the processing of very large data volumes, to integrate into the Cloud or to deploy Big Data applications. BullSequana M is the new range of mainframe class servers with unmatched features. Designed by Atos engineers, based on the latest Intel® Xeon® Scalable processors and a highly flexible and modular architecture, BullSequana M servers benefit from the latest technological features such as virtualization.
4. Bull eXascale Interconnect (BXI): Exascale entails an explosion of performance, of the number of nodes/cores, of data volume and data movement. At such a scale, optimizing the network that is the backbone of the system becomes a major contributor to global performance. The interconnect is very likely going to be a key enabling technology for exascale systems. This is why one of the cornerstones of Bull's exascale program is the development of our own new-generation interconnect. The Bull eXascale Interconnect or BXI introduces a paradigm shift in terms of performance, scalability, efficiency, reliability and quality of service for extreme workloads. The

BXI fabric is highly scalable (up to 64.000 nodes for the first version), it features:

- High-speed links (100 Gb/s/s)
- High message rate (>100 M msg/s)
- Minimal memory footprint and low latency components.

The BXI fabric relies on two types of ASICs as its building blocks, a Network Interface Controller (NIC) and a switch and comes with its complete software suite. BXI switches are managed through a distributed and out-of-band fabric management suite allowing to scale up to 64K nodes. Out-of-band management eliminates any interference of the management traffic with the applications traffic.

5. Bull supercomputer suite, or SCS: Bull supercomputer suite introduces a new approach to extreme computing software solutions. Bull SCS is a scalable, open, and robust software suite that meets the requirements of even the most challenging high-performance computing (HPC) environments, which also require enhanced security. Bull SCS is the result of Atos's long experience in deploying large-scale supercomputers, combined to continued efforts in Research & Development. Bull supercomputer suite is designed for every HPC need, from small supercomputers with just a few hundred cores to supercomputers with tens of thousands of nodes. It is cut out to reach performance targets of the order of up to 100 PFlops, based on new-generation CPUs and GPUs. The main goal of Bull SCS is to provide a global high-performance supercomputing environment. It includes:

- a standardized and scalable installation process with an enhanced update solution; mechanisms to ease integration of new hardware;
- default security with on-time fixes;
- and support for several user development and execution environments with top performance.

This new generation HPC software suite is a further step towards Exascale computing.

Atos BDS has developed Atos Codex AI Suite, which is part of the Atos Codex portfolio. Atos Codex AI Suite is a software suite dedicated to artificial intelligence, adapting to multiple use cases, leveraging the technologies of Big Data, Machine Learning and Deep Learning to enable rapid deployment of software environments for data analysis and artificial intelligence on heterogeneous platforms, whether supercomputer (HPC), enterprise (Big Data) or Cloud. Atos Codex AI leverages years of experience developing HPC and Big Data products to deliver software solutions that accelerate convergence between the worlds of digital simulation, data analytics, and artificial intelligence.

Relevant previous projects/activities

Atos BDS is involved with the strategy toward HPC in Europe with the active leadership of ETP4HPC and contribution to the Strategic Research Agenda.
Atos BDS has been involved in the following finished and ongoing cooperative projects connected to the subject of this proposal.

- EU H2020 Mont-Blanc 2020 project - ambitions to initiate the development of a future low-power European processor for Exascale. MB2020 lays the foundation the European Processor Initiative (EPI)
- EU H2020 EPI - European Processor Initiative aiming to design and implement a roadmap for a new family of low-power European processors for extreme scale computing, high-performance Big-Data and a range of emerging applications
- EU H2020 CloudDBAppliance - European Cloud In-Memory Database Appliance with Predictable Performance for Critical Applications
- EU H2020 LEXIS - Large-scale EXecution for Industry & Society - building an advanced engineering platform to leverage large-scale geographically-distributed resources from existing HPC infrastructure, to employ Big Data analytics solutions and augment them with Cloud services in order to help organizations facing business challenges of seeking to glean knowledge and understanding from the data. The platform will be validated by the pilots in the industry & scientific sectors (Aeronautics, Earthquake and Tsunami, Weather and Climate).
- FR PIA HYDDA - HYbrid Deployment for Data Analytics, is a French national project dealing about big data workflows deployment on hybrid Cloud/HPC infrastructures.

Relevant infrastructures and networks

Atos will use the following platforms for its own testing and experimenting activities in collaborative R&D projects:

- Platform NOVA is a cluster architecture hosted in Atos France facilities providing a small datacenter for Cloud/Big Data as well as HPC/Simulation project. This heterogeneous platform includes a mix of standard x86 servers, large In-Memory Servers (Bullion-S8 et –S16), HPC servers (bullx 515, 520, R423-E4i) with or without accelerators (NVIDIA GPU or INTEL XeonPhi) and also storage systems NetApp-LSI 2800 (240TB) and 2 NetApp 2700 (98TB and 48TB). This platform that can be tuned is providing cloud environments under OpenStack or an HPC environment running an enhanced operating system, lustre file system, development tools and middleware software for executing compute-intensive applications.
- a new Arm-based prototype, called Platform DIBONA to test software and scientific applications on a state-of-the-art Arm architecture. It is based on 64-bit ThunderX2 processors from Marvell®, relying on the Arm® v8 instruction set. The prototype leverages the BullSequana X1000 infrastructure, including ultra-efficient Direct Liquid Cooling – cooling with warm water. The system also features the Mont-Blanc optimized software stack and an InfiniBand EDR interconnect.

4.1.6. DtoK Lab S.r.l.

Short name	DtoK Lab	
Type	SME	
Website	www.dtoklab.com	
Country	Italy	

Description of the organization
DtoK Lab is an academic spin-off of University of Calabria, Italy, working on innovative software solutions for Big Data analysis on Clouds and HPC systems for exploiting customers data. The company is committed to deliver scalable solutions for Big Data analysis in business and scientific domains using Cloud and high-performance technologies. DtoK Lab was awarded by Microsoft and Intel with special prizes at Premio Nazionale per l'Innovazione (PNI) 2013. The company develops Software-as-a-Service (SaaS) systems and workflow-based toolkits for large data analysis applications. DtoK Lab technology makes it possible to parallelize the execution of data-intensive applications exploiting Cloud and HPC storage and processing resources and services. Examples of such systems are DMCF, Nubytics and SMART. Customers of DtoK Lab are enterprises, business analysts, institutions that need to efficiently analyze the huge amount of data collected every day in their data centers or public Clouds. DtoK Lab helps them understand how to uncover and explore those data. DtoK Lab provides consulting and implements high-performance data analysis in emerging areas such as social media, trajectory mining, data science, sentiment analysis and data journalism. DtoK Lab will provide to the project skills and expertise that include research and development know how in data mining, Big Data management and analysis, workflows, cloud computing, high-performance data analysis in emerging scientific areas included in the project proposal.
Role of Partner in the Project
Tasks in the project: **Task 1.2:** Contribution to designing the architecture of the workflows software stack to be integrated in the project with focus on data-driven and service-oriented aspects. **Task 1.3:** DtoK Lab leads this task and provides contribution for designing the interfaces for the integration of HPC, data analytics and machine learning in a single workflow. **Task 2.1:** DtoK Lab leads this task and provides contribution to developing optimizations for a scalable workflow environment with a particular focus on data management issues. **Task 2.5:** Contribution to the design and development of data management strategies in service-oriented workflow deployment on Cloud and HPC systems. **Task 5.2 - Task 5.6:** Contribution to the design, development, validation and integration of results of the Pillar I use cases on top of the eFlows4HPC software stack. **Previous relevant expertise:** The relevant expertise of DtoK Lab is in several research areas of interest for the eFlows4HPC project. In particular DtoK Lab personnel has significant expertise in machine learning techniques, scalable data mining algorithms, scientific workflow management systems, service-oriented architectures, data-intensive Cloud computing systems, parallel programming, Big Data analysis applications. **Expected benefits:** The project will benefit from DtoK Lab participation in terms of research and development skills in some of the key topics such as service-oriented workflow systems, parallel and distributed data analysis and machine learning techniques, Clouds and HPC programming. The benefits will come mainly from scientific and professional background of DtoK Lab specialists who worked on similar research issues and implemented research prototypes and software products in the area of the project. DtoK lab will also provide software systems that can be usefully integrated in the project software stack for handle significant system functionality related to workflows and data analysis.
Qualification of key personnel

Domenico Talia (male) is a partner of DtoK Lab and is a full professor of computer engineering at the University of Calabria and an adjunct professor at Fuzhou University. His research interests include Big Data analysis, parallel and distributed data mining algorithms, Cloud computing, distributed knowledge discovery, mobile computing, distributed computing, peer-to-peer systems, and parallel programming. Talia published ten books and about 400 papers in archival journals and international conference proceedings. He is a member of the editorial boards of IEEE Transactions on Parallel and Distributed Computing, the Future Generation Computer Systems journal, the International Journal on Web and Grid Services, the Scalable Computing: Practice and Experience journal, the Journal of Cloud Computing, and the Web Intelligence and Agent Systems International journal. He served as a program chair and program committee member of several scientific conferences and gave many invited talks and seminars in international conferences and schools. Talia worked in several EU projects. He is a member of the ACM and a senior member of IEEE.

Paolo Trunfio (male) is the managing director of DtoK Lab S.r.l. and an associate professor of computer engineering at DIMES Department, University of Calabria, Italy. He was visiting researcher at the Swedish Institute of Computer Science in Stockholm (2007) and a research collaborator at the Italian National Research Council (2001-2002). Trunfio published three books and many papers in archival journals and international conference proceedings He is currently serving as associate editor of the IEEE Transactions on Cloud Computing and is a member of the editorial board of several scientific journals, including Future Generation Computer Systems, the Journal of Big Data, and the International Journal of Web Information Systems. He is a senior member and a distinguished speaker of the ACM.

Fabrizio Marozzo (male) is a partner of DtoK Lab and is an Assistant Professor at University of Calabria. He received a Master's Degree in computer engineering and a Ph.D. in Systems and Computer Engineering at the University of Calabria. In 2011-2012 he visited the Barcelona SuperComputing Center (Spain) for a research internship with the Grid Computer Research group in Computer Sciences department. His current research focuses on cloud computing, parallel and distributed data analysis, intelligent systems, peer-to-peer networks, social network and big data analysis. He has published more than 50 scientific papers in international conference, proceedings and international journals. He is co-author of a book related to Big Data analysis on distributed computing platforms: "Data Analysis in the Cloud" (Elsevier, 2015).

Francesco Marzano (male) is a full stack developer at DtoK Lab, with expertise on cloud-based solutions for data processing and the major front-end technologies. He obtained a Master of Science in Computer Engineering from University of Calabria in 2017, with a thesis on the optimization of a distributed peer-to-peer protocols for improving energy efficiently. He worked on the SMART project funded by Regione Calabria under EU-FESR research and development program.

Publications/achievements relevant to the call

- D. Talia, P. Trunfio, F. Marozzo, "Data Analysis in the Cloud", Elsevier, 2015.
- L. Belcastro, F. Marozzo, D. Talia, "Programming Models and Systems for Big Data Analysis". International Journal of Parallel, Emergent and Distributed Systems, vol. 34, pp. 632-652, 2019.
- F. Marozzo, D. Talia, P. Trunfio, "A Workflow Management System for Scalable Data Mining on Clouds". IEEE Transactions On Services Computing, vol. 11, n. 3, pp. 480-492, 2018.
- ParSoDA, Parallel Social Data Analytics toolkit, github.com/SCAlabUnical/ParSoDA.
- Nubytics - a Software-as- a-Service (SaaS) system for analyzing large datasets, www.nubytics.com.

Relevant previous projects/activities

Nubytics – In this project has been implemented Nubytics is a Software-as-a-Service (SaaS) system that exploits Cloud facilities to provide efficient services for analyzing large datasets. The system allows users to import their data to the Cloud, extract knowledge models using high performance data mining services, and exploit the inferred knowledge to predict new data and behaviors.

DMCF – The project Data Mining Cloud Framework (DMCF) implemented a system allowing domain experts to design and execute complex data analysis workflows on Cloud platforms, relying on Cloud processing and storage services for every I/O operation.

SMART – The SMART (Social Media AnalizeR Toolkit) project implemented a software toolkit for the analysis of Big Data coming from different sources. The toolkit has been implemented in Java and provide an easy-to-use interface for data extraction and analysis.

ParSoDA – The ParSoDA (Parallel Social Data Analytics) project developed a Java library for developing parallel data analysis applications based on the extraction of useful knowledge from web and social media data. Applications developed with ParSoDA can be executed on Hadoop, Spark, and PhyCOMPSs frameworks.

DEEP – This project implemented the innovative platform DEEP (Data Enrichment for Engaging People), a software platform for Big Data analysis and prediction exploited in the tourism and cultural heritage domains.

Relevant infrastructures and networks

None.

4.1.7. Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici.

Short name	Fondazione CMCC	
Type	Research Centre	
Website	www.cmcc.it	
Country	Italy	

Description of the organization

The Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Fondazione CMCC) is a non-profit research institution. CMCC's mission is to investigate and model our climate system and its interactions with society to provide reliable, rigorous, and timely scientific results, which will in turn stimulate sustainable growth, protect the environment, and develop science driven adaptation and mitigation policies in a changing climate. CMCC collaborates with experienced scientists, economists, and technicians, which work together in order to provide full analyses of climate impacts on various systems such as agriculture, ecosystems, coasts, water resources, health, and economics. CMCC also supports policymakers in setting and assessing costs, mitigation, and adaptation policies. CMCC benefits from the extensive applied research experience of its members and institutional partners: Istituto Nazionale di Geofisica e Vulcanologia (INGV); Università del Salento; Centro Italiano di Ricerche Aerospaziali (CIRA S.c.p.a.); Università Ca' Foscari Venezia; Università di Sassari; Università della Tuscia; Università di Bologna; Politecnico di Milano; Resources for the Future.

CMCC research activities are distributed among nine research divisions that share different knowledge and skills in the field of climate science: Advanced Scientific Computing (ASC) Division; Climate Simulation and Prediction (CSP) Division; Economic analysis of Climate Impacts and Policy (ECIP) Division; Impacts on Agriculture, Forests and Ecosystem Services (IAFES) Division; Ocean modeling and Data Assimilation (ODA) Division; Ocean Predictions and Applications (OPA) Division; Risk Assessment and Adaptation Strategies (RAAS) Division; Regional Models and geo-Hydrological Impacts (REHMI) Division; Sustainable Earth Modeling and Economics (SEME) Division.

CMCC acquired portfolio of research projects includes 358 funded projects: 37 projects in FP6 and FP7, 57 projects in H2020 and 264 projects under other EU and international research grants. In about a half of the implemented projects, CMCC acted as the coordinator.

Role of Partner in the Project

Tasks in the project: CMCC will lead WP5 about the Pillar II on Dynamic and adaptive workflows for climate modelling. CMCC will contribute to all the tasks planned in WP5, leading in particular, the two development phases (Task 5.3 and Task 5.4).

CMCC will also participate in WP1, WP2, WP3, WP7 and WP8, by providing respectively a contribution to the:

- design and development of workflow interfaces for the integration of HPC, data analytics, and ML (WP1);
- optimization of ML/DL libraries, data management, storage technologies and deployment of HPDA modules through containerization over HPC systems (WP2);
- co-design aspects between applications, software stack and hardware, in particular adaptation and optimization of application kernels to heterogeneous components (WP3)
- training activities (WP7) as well as dissemination and communication (WP8).

Previous relevant expertise: CMCC will contribute to this project through the Advanced Scientific Computing (ASC) and the Climate Simulation and Prediction (CSP) Divisions.

The ASC Division of CMCC carries out Research & Development activities on Computational Sciences applied to Climate Change. In particular, the ASC Division works on: (i) the optimization and parallelization of numerical models for climate change simulations; (ii) the development of open source solutions addressing efficient access, analysis, and mining of scientific data, and (iii) data science and learning environments integrating a) workflow tools, b) machine/deeep learning technologies as well as c) big data and analytics frameworks, into HPC/cloud environments to tackle climate change challenges at scale.

CMCC provides a big data analytics framework (Ophidia – http://ophidia.cmcc.it) targeting in-memory, parallel data analysis on scientific/multidimensional data, like Earth Observation and climate change data.

Presently, the ASC CMCC Division is involved into the EU H2020 projects EOSC-HUB, ESiWACE2, EXDCI2, IS-ENES3, EOSC-Pillar and it has been involved into the EU H2020 INDIGO-DataCloud, BARRACUDA (RDA EUROPE3), IS-ENES2, EUBra-BIGSEA projects working on scientific data management, server-side and parallel data analysis, PaaS and workflows for big data analytics, to address scientific case studies.

The CSP Division of CMCC is involved in the investigation of tropical cyclones activity and their relationship with the climate system through the usage of reanalysis and general circulation models since 2007 - providing about 20 contributes to scientific journals - participating to international projects such as the TCMIP endorsed by the WCRP and the US-CLIVAR hurricane working group. CSP has also provided high resolution climate data, to the Coupled Model Intercomparison Projects CMIP3 and CMIP5 and is currently working on the CMIP6 effort also through the provision of very high-resolution climate data within the HigResMIP protocol based on the PRIMAVERA EU H2020 project.

Expected benefits: The improved workflow capabilities provided by eFlows4HPC will contribute to enhance and optimize, climate simulations workflows running at CMCC. eFlow4HPC will help providing a more integrated and holistic "tool" to CMCC users, joining together numerical simulations and HPDA, ML/DL components.

eFlows4HPC will allow the consolidation and further advancement of CMCC expertise/skills in the workflow, HPDA and AI scientific areas. From an operational standpoint, data analytics production workflows running at the center will leverage and benefit from the novel capabilities and support offered by eFlow4HPC. The development and optimization of workflows to extract added-value products like Tropical Cyclones Tracks will help CMCC scientists working in this area to enhance their scientific research and ability to deal with larger datasets/ensembles, as well as more complex workflows. Finally, the eFlows4HPC software stack will be deployed on the CMCC Zeus Super Computing system to better support scientists with respect to HPDA and ML/DL over HPC as well as workflow support.

Qualification of key personnel

Giovanni Aloisio (male) is Full Professor of Information Processing Systems at the Dept. of Innovation Engineering of the University of Salento, Lecce, Italy, where he leads the HPC laboratory. He was the director of the "Advanced Scientific Computing" (ASC) Division at the Euro-Mediterranean Center on Climate Change (CMCC) from 2005 to 2014. He has been the director of the CMCC Supercomputing Center since it was first created. He was a member of the CMCC Board of Directors and the CMCC Executive Committee from 2015 to 2017. From 2014, he has been a member of the CMCC Strategic Council. His expertise concerns high performance computing, grid & cloud & exascale computing and distributed data management. He has been involved in several EU projects, such as GridLab, EGEE, IS-ENES, EESI, EXDCI, EUBrazilCC, INDIGO-DataCloud, OFIDIA, EUBra-BIGSEA. He is currently involved in the H2020 European projects IS-ENES3, ESCAPE-2, IMMERSE, ESiWACE-2, EXDCI-2, OFIDIA-2. He has also contributed to the IESP (International Exascale Software Project) exascale roadmap. He has been the chair of the European panel of experts on "Weather, Climate & solid Earth Sciences" (WCES), that has contributed to the PRACE strategic document "The Scientific Case for HPC in Europe 2015-2020". He is a member of the HPC Task Force of the "European Network for Earth System Modeling" (ENES) and part of the Scientific Steering Committee of the Nemo Consortium. He is the author of more than 150 papers in refereed journals on high performance computing, grid & cloud & exascale computing and distributed data management.

Dott. Sandro Fiore, Ph.D., (male) Data Scientist and Leader of the Data Science and Learning Research Team at the Euro-Mediterranean Center for Climate Change (CMCC) Foundation. His research activities focus on High Performance and Distributed Computing, with specific regard to distributed data management, big data analytics, high performance database management, data mining and machine learning. In 2006, he joined as data scientist the CMCC Scientific Computing Division to lead the Scientific Data Management Research Group. From November 2014 to October 2018, he has been the Director of the Advanced Scientific Computing Division. He has been involved in several EU Projects such as IS-ENES, EUBRAZILCC, CLIP-C, EUBra-BIGSEA, INDIGO-DataCloud, EESI, EXDCI and he is currently

involved in EXDCI2, EOSC-Hub and ESiWACE. Since 2010, he has been the Principal Investigator of Ophidia, a research project on high performance data analytics and mining for eScience. Since 2011, he has been Visiting Scientist at the Lawrence Livermore National Laboratory. Since 2019, he has been Visiting Scientific at the University of Chicago. He is a member of the ENES Data Task Force. He is co-author of about 100 papers in refereed books, journals and proceedings on parallel, distributed and grid computing, and he holds a patent on data management topics. He is an ACM Member, editor of the book "Grid and Cloud Database Management" (Springer, 2011) and co-author of The International Exascale Software Project roadmap, IJHPCA 25(1), pp. 3-60 (2011).

Donatello Elia (male) is a Junior Data Scientist for the Advanced Scientific Computing (ASC) division at the Euro-Mediterranean Center on Climate Change (CMCC) Foundation, which he joined in 2013. He holds a M.Sc. degree in Computer Engineering (2013) from the University of Salento in Italy, where he is currently pursuing a PhD degree in Engineering of Complex Systems. His main research interests include high performance and distributed computing, data-intensive analytics, big data management and data mining. Since 2018, he has been the Co-P.I. of the Ophidia project, a research effort on High Performance Data Analytics for eScience. He has been involved in various European projects, like the FP7 and Horizon 2020 programmes. He has also authored and co-authored various papers in refereed journals and conference proceedings.

Eng. Alessandro D'Anca, (male) is a Scientist at CMCC Foundation and Director of the Advanced Scientific Computing (ASC) Division. He earned a degree in Computer Engineering at the University of Lecce in 2006. From 2008 to 2011, he worked as Junior Computer Systems Analyst for the ComputerVar S.r.l company at the Euro-Mediterranean Center on Climate Change (CMCC) in Lecce. He joined the Scientific Computing and Operations Division (current ASC division) at CMCC in 2011, where he started working in the Scientific Data Management research group. In 2015, he became Head of the Scientific Data Management research group of the ASC Division and was later appointed Division Director in 2018. His research activities focus on high performance computing, distributed and grid computing, and in particular on distributed data management, data analytics/mining and high-performance database management. He has been involved in many national and international projects (TESSA, CLIP-C, OFIDIA, MARSOP4, INDIGO-Datacloud, ESIWACE), working on development, project or scientific management and coordination tasks. He is also the author and co-author of several papers on big data analytics for eScience.

Enrico Scoccimarro (male) is senior scientist at CMCC Foundation, within the Climate Simulations and Predictions division. He has contributed to the development of several fully coupled climate models (e.g., SINTEX-G, CMCC-Med, CMCC-CM, CMCC-CM2) and has performed a number of climate scenario simulations (CMIP3, CMIP5, CMIP6). He is an expert of extreme events and his main research activity is to investigate the relationship between Tropical Cyclones and Climate. He has been a member of the Tropical Cyclone Model Intercomparison Project (TCMIP) and the US-CLIVAR Hurricane Working Group since 2011. Enrico participated to many EU research projects, also as CMCC Principal Investigator or Work Package leader, and he is author of more than 50 papers on scientific journals.

Silvio Gualdi (male) is the Director of the "Climate Simulation and Prediction" (CSP) Division at CMCC, Italy. MSc in Physics and PhD in Geophysics. He has more than 20 years of experience in climate modelling and simulations. He is expert in simulation and analysis of climate variability from intra-seasonal to multi-decadal timescales and climate change, through: i) development, implementation and evaluation of coupled GCMs used to produce simulations of past, present and future climate; ii) production of climate predictions (seasonal-to-multi annual) and climate change projections; iii) investigation and assessment of the interaction and teleconnection mechanisms, respectively, between the major modes of tropical variability at different timescales, e.g. El Nino, NAO, etc. and between the tropics and the Mediterranean region.

Publications/achievements relevant to the call

- S. Fiore, D. Elia, C. E. S. Pires, D. G. Mestre, C. Cappiello, M. Vitali, N. Andrade, T. Braz, D. Lezzi, R. Moraes, T. Basso, N. P. Kozievitch, K. V. Ono Fonseca, N. Antunes, M. Vieira, C. Palazzo, I. Blanquer, W. Meira Jr., G. Aloisio, "An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management", in IEEE Access, vol. 7, pp. 117652-117677, 2019. doi: 10.1109/ACCESS.2019.2936941.

- A. S. Alic, J. Almeida, G. Aloisio, N. Andrade, N. Antunes, D. Ardagna, R. M. Badia, T. Basso, I. Blanquer, T. Braz, A. Brito, D. Elia, S. Fiore, D. Guedes, M. Lattuada, D. Lezzi, M. Maciel, W. Meira, D. Mestre, R. Moraes, F. Morais, C. E. Pires, N. P. Kozievitch, W. dos Santos, P. Silva, M. Vieira, "BIGSEA: A Big Data analytics platform for public transportation information", Future Generation Computer Systems, Volume 96,2019, Pages 243-269, ISSN 0167-739X, https://doi.org/10.1016/j.future.2019.02.011.
- S. Fiore, D. Elia, C. Palazzo, F. Antonio, A. D'Anca, I. Foster, G. Aloisio (2019), "Towards High Performance Data Analytics for Climate Change". In: Weiland M., Juckeland G., Alam S., Jagode H. (eds) High Performance Computing. ISC High Performance 2019. Lecture Notes in Computer Science, vol 11887. Springer, Cham, https://doi.org/10.1007/978-3-030-34356-9_20.
- D. Salomoni, I. Campos, L. Gaido, J. Marco de Lucas, P. Solagna, J. Gomes, L. Matyska, P. Fuhrman, M. Hardt, G. Donvito, L. Dutka, M. Plociennik, R. Barbera, I. Blanquer, A. Ceccanti, E. Cetinic, M. David, C. Duma, A. López-García, G. Moltó, P. Orviz, Z. Sustr, M. Viljoen, F. Aguilar, L. Alves, M. Antonacci, L. A. Antonelli, S. Bagnasco, A. M. J. J. Bonvin, R. Bruno, Y. Chen, A. Costa, D. Davidovic, B. Ertl, M. Fargetta, S. Fiore, S. Gallozzi, Z. Kurkcuoglu, L. Lloret, J. Martins, A. Nuzzo, P. Nassisi, C. Palazzo, J. Pina, E. Sciacca, D. Spiga, M. Tangaro, M. Urbaniak, S. Vallero, B. Wegh, V. Zaccolo, F. Zambelli, T. Zok, "INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures", Journal of Grid Computing (2018) 16: 381. https://doi.org/10.1007/s10723-018-9453-3.
- Scoccimarro E., P.G. Fogli. K. Reed, S. Gualdi, S.Masina, A. Navarra: "Tropical cyclone interaction with the ocean: the role of high frequency (sub-daily) coupled processes". Journal of Climate, doi: 10.1175/JCLI-D-16-0292.1

Relevant previous projects/activities

EU H2020 Project **ESIWACE2** (Centre of Excellence in Simulation of Weather and Climate in Europe; 2015-2019). ESIWACE is a user-driven Centre of Excellence in Simulation of Climate and Weather in Europe. In particular, CMCC Foundation contributes to enhancing community capacity in HPC, scheduling and workflow capabilities, as well as storage, processing and analytics middleware.

EU FP7 Project IS-ENES (2009-2013), IS-ENES2 (2013-2017) and **H2020 IS-ENES3** (2019-2022). The goal of IS-ENES is the development of a common climate and Earth system modelling distributed research infrastructure in Europe. CMCC coordinates the development of the ENES Climate Data Infrastructure as well as it participates in a number of activities from data management to model development to technology tracking, networking and dissemination.

EU H2020 Project **INDIGO-DataCloud** (INtegrating Distributed data Infrastructures for Global ExplOitation; 2015-2017). INDIGO-DataCloud has focused on the development of a data/computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over hybrid (private or public) e-infrastructures. CMCC mainly contributed to data analytics workflow support for ENES and other scientific communities.

EU H2020 **EOSC-Hub** (Integrating and managing services for the European Open Science Cloud; 2018-2020). The EOSC-hub project creates the integration and management system of the future European Open Science Cloud that delivers a catalogue of services, software and data from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-infrastructures. In the project, CMCC co-leads the ECAS (ENES Climate Analytics Service) Thematic Service activities for the climate community.

EU H2020 Project **PRIMAVERA** (PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment; 2015-2019). The goal of PRIMAVERA is to deliver novel, advanced and well-evaluated high-resolution global climate models (GCMs), capable of simulating and projecting regional climate with unprecedented fidelity, out to 2050. The CSP Division participated as WP leader and also developed the CMCC-CM2-HR and -VHR models, providing the relative scenario simulations. The ASC division of the CMCC Foundation contributes to WP9 addressing HPC and Data management challenges.

Relevant infrastructures and networks

Housed in the Ecotekne Campus of the University of Salento (Lecce), the CMCC Supercomputing Center was inaugurated in January 2009. It provides the technological infrastructure and the computational capabilities needed in order to create and develop scenarios and models of the climate of the future. The IT facilities of the CMCC Supercomputing Center consists of two parallel supercomputers called Athena and Zeus with their own HPC storage systems; a Tier2 level storage systems: a backup and archiving system: tape library and related management systems.

The Athena supercomputer was purchased in 2012 and it has been in production at SCC since February 2013. Athena is a parallel scalar computing system based on 482 IBM iDataPlex dx360 M4 compute nodes (for a total of 7712 cores) all connected to a FDR Infiniband network and has a computing capability of 160TFlops. The HPC storage infrastructure connected to Athena consists of two DDN SFA10000 storage systems, with identical configuration, which offer, overall, a usable capacity of 600 TeraBytes.

The Zeus parallel scalar supercomputer was installed in the first half of 2019. Zeus is based on 348 Lenovo SD530 biprocessor nodes (for a total of 12.528 cores) all interconnected by means of an Infiniband EDR network. The new system has a computing power (theoretical peak performance) of 1.2 PetaFlops. The HPC storage infrastructure connected to Zeus consists of two Lenovo DSS-G260 storage systems, with identical configuration, which offer, overall, a usable capacity of 4 PetaBytes.

During 2019, the data center was also equipped with a level 2 (Tier2) storage system which offers a total capacity of over 4 PetaBytes. This system belongs to the storage class that meets the requirements of large storage capacity, instant data access (online storage), high levels of scalability and protection and is oriented to the provision of services, with a high degree of independence from data processing infrastructures and connected applications. In order to store and manage the huge amount of data produced by CMCC researchers, the data center is equipped with an archiving system (tape library) with a usable capacity of 5 PetaBytes.

4.1.8. Institut National de Recherche en Informatique et Automatique

Short name	INRIA	
Type	Research Center	
Website	www.inria.fr	
Country	France	

Description of the organization

INRIA, the French national research institute for the digital sciences, promotes scientific excellence and technology transfer to maximise its impact. It employs 2,400 people. Its 200 agile project teams, generally with academic partners, involve more than 3,000 scientists in meeting the challenges of computer science and mathematics, often at the interface of other disciplines. Inria works with many companies and has assisted in the creation of over 160 startups. It strives to meet the challenges of the digital transformation of science, society and the economy

INRIA has been very active in the previous European framework programmes (in FP7: 231 projects).

The institute is strongly involved in programmes aimed at fostering scientific excellence, such as the European Research Council: 32 Grants in FP7 and 23 Grants in Horizon 2020.

INRIA makes a firm commitment to Horizon 2020, with which the institute's strategic plan is aligned. The objective is to combine scientific excellence with a more focused consideration of major European and global societal challenges to which Inria can bring a key contribution. Inria is currently involved in more than 120 H2020 funded projects.

INRIA is playing a lead role in the construction of the Knowledge and Innovation Community (KIC) EIT Digital as host of the French node. EIT Digital's ambition is to create for Europe a structure dedicated to technology transfer and innovation in the digital field. Besides EIT Digital, Inria is a core partner of the KIC EIT Health. More information: http://www.inria.fr

Role of Partner in the Project

INRIA will contribute to the "Manufacturing Pillar" within which will provide leading expertise in the development of high-performance adaptive solvers and mesh adaptation tools, including the MMG and ParMMG platforms. The project team includes in-house experts in the field of immersed and embedded methods, of error estimation and anisotropic mesh adaptation, as well as of the interaction of the two.

Within the pillar, the INRIA team will work towards the definition of a Digital Twin to be used both in the design phase and as a life companion of manufactured products. The main goal will be the development of adaptive techniques exploiting the use of large-scale HPC hardware and the interaction of Reduced Order Modelling and adaptation to provide faster and more accurate digital representations.

One of the main research interests will be the integration of the mesh adaptation capabilities available in the group, with high order embedded boundary models for manufacturing, and with reduced order models of the geometrical error, within BSCs workflow management systems. More particularly, the team will benefit from the experience gained in the study of the interaction and integration of Machine Learning and of Artificial Intelligence techniques with geometrical error estimation and adaptation.

Qualification of key personnel

Nicolas Barral: PhD at Inria (GAMMA3 team) and University Paris 6 Pierre et Marie Curie, and Research Assistant in the Department of Earth Science and Engineering at Imperial College London. Joined Bordeaux INP as an Assistant Professor and the Inria CARDAMOM team in September 2019. Expert in mesh adaptation in an HPC context for various applications going from three-dimensional unsteady compressible flow simulations tor geophysical flows. Best Paper Award at the AIAA 2019 SciTech Forum and Exposition for the paper "Verification of Unstructured Grid Adaptation Components".

Heloise Beaugendre: PhD from McGill University (Montreal, Canada), Assistant Professor and and Head of the High Performance Computing track of the Bordeaux INP engineering school. Heloise Beaugenre is an expert in scientific computing, and more particularly in the numerical approximation of partial differential equations. Her main field of application is the study, modeling, and simulation of wing icing systems for which she has investigated different numerical techniques going from particle in cell methods to finite element, volume, and residual distribution schemes.

Algiane Froehly: Research engineer in applied mathematics, specialized in meshing tools and software design. After a PhD thesis on high order iso-geometric numerical methods and high order curved simplicial mesh generation, I started to work on the industrialization of the Mmg re-meshing platform. I am one of the main contributors to the construction of the Mmg open source consortium gathering academics and industrial members steering and endorsing the development the Mmg tools. Currently, I am in charge of the consortium management, of the promotion and distribution of the platform, as well as and of its technical development.

Mario Ricchiuto: scientific leader of the Inria team CARDAMOM and co-lead of the associated team HAMSTER, joint with the CEE department at Duke University (co-lead Prof. G. Scovazzi), appointed Adjunct professor at the CEE depatment at Duke in Novemner 2019. Mario Ricchiuto obtained his PhD at the von Karman Institute for Fluid Dynamics and Université Libre de Bruxelles, and is an expert in the numerical approximation of nonlinear partial differential equations on adaptive unstructured grids. He has contributed to 43 articles in major journals in the computational mechanics and physics, applied mathematics and fluid dynamics communities, as well as 18 chapters in books and lecture series, including two in the Encyclopedia of Computational Mechanics. He is member of the editorial boards of Computers & Fluids (Elsevier), and of Water Waves (Springer), and of the CFD scientific committee of the 14th World Conference on Computational Mechanics. He has coordinated or represented Inria in several National and International funding programs.

Publications/achievements relevant to the call

- MMG tools: open source software for simplicial re-meshing allowing surface and volume (adaptive) mesh improvement in serial and parallel. Distributed under LGPL licence: www.mmgtools.org

- L. Nouveau, M. Ricchiuto, G. Scovazzi, "High-Order Gradients with the Shifted Boundary Method: Aa Embedded Enriched Mixed Formulation for Elliptic PDEs", J. Comput. Phys., doi: 10.1016/j.jcp.2019.108898

- N. Barral, G. Olivier and F. Alauzet, "Time-accurate anisotropic mesh adaptation for three-dimensional tme-dependent problems with body-fitted moving geometries", Journal of Computational Physics, 2017

- R. Abgrall, H. Beaugendre, and C. Dobrzynski, "An immersed boundary method using unstructured anisotropic mesh adaptation combined with level-sets and penalization techniques", J. Comput. Phys. 257, 2014

- Ch. Dapogny, C. Dobrzynski and P. Frey, "Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems", J. Comput. Phys, 262, 2014

Relevant previous projects/activities

European project H2020-EU.1.2.2 ExaQUte (started 2018): 3-year collaborative project aiming at constructing a framework to enable Uncertainty Quantification (UQ) and Optimization Under Uncertainties (OUU) in complex engineering problems using computational simulations on Exascale systems. The stochastic problem of quantifying uncertainties will be tackled by using a Multi Level MonteCarlo (MLMC) approach that allows a high number of stochastic variables. New theoretical developments will be carried out to enable its combination with adaptive mesh refinement, considering both, octree-based and anisotropic mesh adaptation. All developments in ExaQUte will be open-source and will follow a modular approach, thus, maximizing future impact. Role: participant.

National Research Agancy (ANR) project VISCAP: 4-years collaborative project on the experimental study as well modelling and simulation of the lifetime of self-healing composite materials under

mechanical and thermal loads. VISCAP aims at creating a true computational platform describing the multi-scale, multidimensional and multi-physics character of the phenomena that determine the material lifetime. Important outcomes in the domain of civil aircraft jet propulsion are expected, that could relate to other materials than those considered in this study. Role: participant.

European OceanERANET MIDWEST (ended 2018): 3-year collaborative project evaluating novel nonlinear multi-fidelity (including surrogate reduced models) modelling approaches for the hydrodynamic optimization of wave energy devices. Role: coordinator.

European project FP7-TRANSPORT STORM (ended 2017) :3-year collaborative project on the study of wing icing and de-anti-icing systems comprising 14 research and industrial partners from 7 European countries. STORM project evaluated new advanced simulation methodologies in three specific themes in the field of wing icing: ice release, ice accretion with runback aspects and ice trajectory applied for aero propulsive systems to improve the knowledge of engine components behaviour under icing conditions. STORM was identified by the WEZARD CSA as a priority research theme within the European R&D roadmap on actions against hazard weather conditions. STORM is also supported by Engines Industries Management Group (EIMG) cluster. Role: participant.

Relevant infrastructures and networks

INRIA manages one of the local large computational clusters allowing experimental runs on modern heterogeneous architectures (PlaFRIM, https://www.plafrim.fr/en/home/), and has access to other computational resources including the new meso-center of the Region Nouvelle Aquitaine https://redmine.mcia.fr/projects/cluster-curta/wiki

4.1.9. Scuola Internazionale Superiore Di Studi Avanzati Di Trieste.

Short name	SISSA	
Type	University	
Website	www.sissa.it	
Country	Italy	

Description of the organization

SISSA, the International School for Advanced Studies, was founded in 1978 and is a scientific center of excellence within the national and international academic scene. Located in Italy, in the city of Trieste, it features 80 professors, about 100 post-docs, 300 PhD students and 100 technical-administrative staff. Situated on the scenic Karst upland, the School is surrounded by a 25-acre park, and offers a stunning view of the Gulf of Trieste. The three main research areas of SISSA are Physics, Neuroscience and Mathematics. All the scientific work carried out by SISSA researchers is published regularly in leading international journals with a high impact factor, and frequently in the most prestigious scientific journals such as Nature and Science. The School has also drawn up over 300 collaboration agreements with the world's leading schools and research institutes. The quality level of the research is further confirmed by the fact that within the competitive field of European funding schemes SISSA holds the top position among Italian scientific institutes in terms of research grants obtained in relation to the number of researchers and professors. SISSA hosts 22 ERC research projects.
The SISSA mathLab group and specifically the group of Prof. Gianluigi Rozza has already active collaborations for the development of methodological research focused on reduced order modelling and uncertainty quantification for Complex Industrial applications together with important partners at the national and international level (Politecnico di Milano, University of Ghent, Imperial College London, Fincantieri, Danieli, Wärtsilä). SISSA is operating an important cluster (Ulysses, 200T Flops).

Role of Partner in the Project

Tasks in the project: SISSA will contribute to all the tasks of the "Manufacturing Pillar" (WP4) of the proposal which is led by the CIMNE team. In particular the SISSA team will lead task 4.3 which is focused on the computational reduction of the test case identified jointly with the other partners.

Previous relevant expertise: The SISSA mathLab team and particularly the group of Prof. Gianluigi Rozza will provide a leading expertise on numerical methods for the development of reduced order methods and digital twins. The group has already been involved in several research projects at both the international (ERC AROMA-CFD, H2020 ARIA) and national level (FARExAROMA-CFD, PRIN national project) focused on the development of novel methods for model order reduction.

Expected benefits: The team with his extensive expertise on scientific computing and especially on computational fluid dynamics, reduced order modelling, data-driven functional approximation will contribute to the development of methods to reduce the computational cost of problems in manufacturing. Such computational reduction is in fact crucial in order to create efficient digital twins.

Qualification of key personnel

Prof. Gianluigi Rozza (male). MS (Politecnico di Milano, 2002). PhD (École Polytechnique Fédérale de Lausanne, 2006). PostDoc (Massachusetts Institute of Technology, 2006-2008). Researcher (École Polytechnique Fédérale de Lausanne, 2008-2012). Currently: Full Professor in Numerical Analysis and Scientific Computing at SISSA. Director's delegate for Innovation and Valorisation at SISSA. Experience in reduced order modelling, uncertainty quantification, numerical analysis, scientific computing, computational mechanics. PI ERC AROMA-CFD, participant in H2020 EID ROMSOC and RISE ARIA, national coordinator PRIN project 2019-2022. Coordinator of several industrial projects with Fincantieri, Danieli, Electrolux Professional, Cetena.

Dr. Andrea Lario (male). MS (Politecnico di Torino, 2011). PhD (Politecnico di Torino, 2017). PostDoc (SISSA, 2018 - present). Responsible for the development of ITHACA-DG, a scientific computing library for model reduction. Experience in Discontinuous Galerkin Method and Reduced Order Models.

Participation in national research program: FARExAROMA-CFD.
Dr. Giovanni Stabile (male). MS (University of Florence, 2012). PhD (TU-Braunschweig - University of Florence, 2016). PostDoc (International School for Advanced Studies, 2016 - present). Responsible for the development of ITHACA-FV a scientific computing library for model reduction. Experience in reduced order modelling for fluid dynamics problems and uncertainty quantification. Participation in national and international R&D programmes: AROMA-CFD. Experience on reduced order models for industrial applications.

Publications/achievements relevant to the call

Rozza, G., Huynh, D. B. P., & Patera, A. T. (2007). "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations". Archives of Computational Methods in Engineering, 15(3), 1.

Hesthaven, J. S., Rozza, G., & Stamm, B. (2016). "Certified reduced basis methods for parametrized partial differential equations", (pp. 27-43). Berlin: Springer.

Stabile, G., & Rozza, G. (2018). "Finite volume POD-Galerkin stabilised reduced order methods for the parametrised incompressible Navier–Stokes equations". Computers & Fluids, 173, 273-284.

Stabile, G., & Rozza, G. (2018). "ITHACA-FV-In real Time Highly Advanced Computational Applications for Finite Volumes". URL http://www. mathlab. sissa. it/ithaca-fv. Accessed, 25-11-2019.

Karatzas, E. N., Stabile, G., Nouveau, L., Scovazzi, G., & Rozza, G. (2019). "A reduced basis approach for PDEs on parametrized geometries based on the shifted boundary finite element method and application to a Stokes flow". Computer Methods in Applied Mechanics and Engineering, 347, 568-587.

Relevant previous projects/activities

Collaboration with the Duke University team of G. Scovazzi on reduced order methods for embedded finite elements.
Collaboration with industrial partners on the development of digital twins and computational pipelines
Prof. Gianluigi Rozza is PI of the ERC-Project AROMA-CFD on Advanced Reduced Order Methods with Applications in Computational Fluid Dynamics (PI Prof. G. Rozza, GA 681447, 2016-2021), Horizon 2020 and local PI for H2020 EID ROMSOC (Reduced Order Modelling for Simulation, Optimization and Control).

Relevant infrastructures and networks

SISSA is operating an important cluster (Ulysses, 200T Flops).

4.1.10. Instytut Chemii Bioorganicznej PAN.

Short name	PSNC	
Type	Research Centre	
Website	www.man.poznan.pl	
Country	Poland	

Description of the organization

Poznań Supercomputing and Networking Center (PSNC) is affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences. PSNC serves as an academic HPC centre and a broadband network services provider as well as application and services developer and provider. PSNC employs about 300 people in four divisions: HPC and data infrastructure, networking, applications and networking services. PSNC is a HPC Center, Systems and Network Security Center as well as R&D Center of Future Internet, e-Infrastructure, Digital Content. It has an active computer science research group working e.g. on: middleware, tools and methods for HPC and distributed computing, resource management, scheduling, large scale applications, user management and accounting, infrastructure security mechanisms and policies, energy aware scheduling, grid and cloud management tools, HPC and distributed storage architectures. PSNC is also the operator of Polish National Research and Education Network PIONIER, which is connected to the GEANT2 and GLIF networks, Operator of Poznań Metropolitan Area Network - POZMAN. PSNC is a leading HPC center in Poland with its 1,7 Pflops of computing power, 42 PB of online storage and data management infrastructure, providing direct support for the scientific communities in Poland as well as in Europe (e.g. Nuclear Fusion, Astrophysics, Bioinformatics, Chemistry, Nanotechnology). In particular PSNC has been for many years strongly involved in supporting the Fusion community through different initiatives like EUFORIA, EGI_Inspire, EURATOM and Eurofusion (nowadays in H2020). It has been coordinating 12 international EU projects (e.g. DORII, RINGRID, GRIDLAB, Phosphorus, Porta Optica, CoolAmAll) and participated in 160+ European and national projects: PRACE, EUDAT/EUDAT2020, IGE, NEXPRES, EGI_INSPIRE, EOSC-Hub and DEEP HYBRID. PSNC is the Microsoft Innovation Center, CUDA Research Center with NVidia and official Cisco Networking Regional Academy. PSNC is one of the founding members of PRACE AISBL, and a partner of Polish NGI. PSNC takes an active part in many international conferences and forums, including the Open Grid Forum, eIRG, TERENA and Geant. PSNC holds the chairman positions in TERENA TF-Storage working group and eIrg Data Management Task Force. PSNC is part of the European HPC infrastructure (PRACE), European and national grid infrastructure (EGI, PL-GRID) as well as national (PLATON e-Science Platform, National Data Storage) and European Data Infrastructure CDI (EUDAT CDI consortium).

Role of Partner in the Project

Tasks in the project: Task 6.2 Design of the Pillar III use cases. Task 6.7 Access policy for urgent computing at HPC infrastructures.

Previous relevant expertise: PSNC is the founding member of PRACE AISBL and continuously takes part in the PRACE Implementation Phase Projects. PSNC is the operator of Tier-1 machine in PRACE. It has the strong expertise in preparing the policies for Urgent Computing scenarios and preparing the piloting environment for urgent modes.

Expected benefits: PSNC expects the tightening the cooperation with the Urgent Computing application providers and its users. To develop such kind of service, there is a need to better understand the requirements coming from parties: crisis decision body, application operator, data providers. Basing on the agreement, PSNC will create the pilot environment for selected eFlows4HPC scenarios. There will be proposed the general template of such co-operation and it will be documented for further use within the project and European Infrastructure accordingly.

Qualification of key personnel

Dr Norbert Meyer (male) is the head of Computing Department in PSNC, responsible for HPC, grid, cloud and data infrastructure management, security provisioning and R&D in the mentioned areas. His research

interests concern resource management in distributed environments, accounting, data management, technology of development graphical user interfaces and network security, mainly in the aspects of connecting independent, geographically distant domains. The concept of remote operation, controlling and monitoring instrumentation of one of key research topics he is currently performing on national and international levels. He is the author and co-author of 80+ conference papers and articles in international journals, member of programme committees of international ICT conferences. Norbert Meyer is the Polish delegate of e-IRG, chair of the data management task force, co-author of several white papers and the main editor of the Blue Paper on Data Management, member of STRATOS group. He coordinated EU projects: DORII and RINGRID, has been participating in project proposals of 5,6 and 7 FP: CrossGrid, EGEE, CoreGRID, Expres, Phosphorus, int.eu.grid, NEXPRES, PRACE projects, EUDAT, IGE, etc. (in total 40+ int. proposals, 2000-2017), coordinator of national projects: MAN-HA, KMD, KMD2, technical coordination - ZPT.

Miroslaw Kupczyk (male), senior IT expert in HPC Department of PSNC. His research interests concern HPC and grid computational environments, resource management, graph algorithms. He has been responsible for putting into practice load sharing facilities (queuing systems) on PSNC production machines. He has prepared and conducted technical tutorials for local and international users concerned distributed processing. He has been involved in several projects (BalticGrid, CrossGrid, EGEE/EGI, FaultTolerant Systems done with France Telecom, PL-Grid, CYBELE, PRACE Implementation projects till now). He is an author and co-author of several reports and papers in scientific journals and conference proceedings. His scientific and engineering interest focuses on HPC processing and large-scale application enabling. He is involved in designing of Urgent Computing service in PRACE RI.

Radosław Januszewski (male), he is an employee at PSNC where he works as leader of HPC and Data Centre Technologies department. He took part in the preparations of some project proposals (e.g. National Data Store, accepted by the Polish Government) and in the design and development of Automatic Backup System for Poznań Metropolitan Area Network. He was involved in work focused on fault tolerance in HTC and GRID environments by means of low-level checkpointing and migration. He is co-author of "Resources virtualization in fault-tolerance and migration issues" paper that was published in ICCS204 proceedings. In years 2001-2004 he worked on designing and implementation checkpoint/migration service for Solaris OS running on SPARC processors. This work was done within the PROGRESS project (http://progress.psnc.pl). He designed and implemented kernel-level checkpointing mechanism for Linux OS running on Itanium 2 CPUs. This effort was done within the confines of the SGIGrid project (http://www.wcss.wroc.pl/pb/sgigrid/en/index.php). 2005-2007 he worked on integration of low-level checkpointing services with the Grid environment for the fault-tolerance purposes. He is co-author of several papers related to fault tolerance in the Grids. In the years 2008-2015 was involved in the PRACE project focusing on HPC energy efficient systems and warm-water, directly liquid cooled HPC solutions. Some of the results were published in paper "Evaluation of the impact of direct warm-water cooling of the HPC servers on the data center ecosystem" in proceedings of the ISC 2014.

Publications/achievements relevant to the call

M. Kupczyk, D. Kaliszan, V. Bergeaud, "Improvement of the Uncertainty Analysis Platform URANIE for High Performance Computing", on-line www.prace-ri.eu, 2014.

Davoli, F.; Meyer, N.; Pugliese, R.; Zappatore, S. (Eds.), "Remote Instrumentation Services on the eInfrastructure", 2011, ISBN 978-1-4419-5573-9, Springer Verlag.

M. Kupczyk, et. al., "PRACE Research Infrastructure Offer for Polish R\&D Community", TASK QUARTERLY vol. 22, No 4, 2018, pp. 285–302, doi:10.17466/tq2018/22.4/f.

Miroslaw Kupczyk. (2017, June 7). "Urgent Computing service in the PRACE Research Infrastructure". Zenodo. http://doi.org/10.5281/zenodo.832029.

Salomoni, Davide & Campos Plasencia, Isabel & Gaido, Luciano & Lucas, J. & Solagna, P. & Gomes, Jose & Matyska, L. & Fuhrmann, Patrick & Hardt, M. & Donvito, Giacinto & Dutka, Lukasz & Plociennik, M.

& Barbera, Roberto & Blanquer, Ignacio & Ceccanti, Andrea & Cetinic, Eva & David, Mario & Duma, C. & Lopez Garcia, Alvaro & Zok, Tomasz. (2018). "INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures. Journal of Grid Computing". 16. 10.1007/s10723-018-9453-3.

Relevant previous projects/activities

Participation to HPC Centres of Excellence: COEGSS, EoCoE.

Participation to former PRACE projects: PRACE /PRACE-1/2/3/4/5/6IP.

Participation to infrastructure related projects funded in FP5,6,7 and H2020 programmes: CrossGrid, EGEE, EGEE2, CoreGRID, BalticGRID, Expres, RinGRID, Phosphorus, int.eu.grid, DORII, NEXPRES, EUDAT and EUDAT2020, EGI.

Participation to EOSC-HUB (Integrating and managing services for the European Open Science Cloud) and DEEP HYBRID – DataCloud.

Participation to national e-infrastructure projects: KMD (National Data Storage), 100Net (New Generation Optical Network, http://www.100net.pionier.net.pl/), Platon (Science Service Platform, http://www.platon.pionier.net.pl/online/), MAN-HA (Deploying high-availability, critical services in Metropolitan Area Networks, http://man-ha.pionier.net.pl/pl/).

Relevant infrastructures and networks

- Tier-1 infrastructure – 1,7 PFlops HPC system, incl. 33+ kcores, 120,6 TB of RAM, Intel Xeon E5 E5-2697 v3 cpus.
- Data infrastructures: 42 PB (archivisation, storage, fast cache memory)
- Operator of the national fibre optic cable network PIONIER (400/100 Gbps technology)
- Operator of the international link to GEANT and GLIF networks.

4.1.11. Universidad de Málaga

Short name	UMA	
Type	University	
Website	www.uma.es	
Country	Spain	

Description of the organization

The University of Málaga (UMA) is a public university established in 1972. With more than 35,000 students and 2,400 lecturers, the UMA offers 62 degree courses, 22 doctoral programs, more than 60 master's degrees, and more than 80 internal degrees and 100 courses throughout the academic year. Teaching takes place in 26 centers by appointed lecturers in 81 departments distributed over two campuses.
About 300 research groups are actively involved in national and international research projects, and in public-private collaborative contracts. While patent portfolio includes about 240 active patents; up to 500 research agreements and contracts are yearly signed with a variety of business and private organization. So far, UMA has participated in more than 200 international research projects, among which 48 were funded by FP7, 37 by Horizon 2020, and up to now, 10 projects have been coordinated.
Our main strengths in EU R&D projects are in the field of the ICTs, Agriculture and Environment related topics, concentrating 46% of all the EU funded projects, followed by Environmental Sciences (12%), Socio-economic Sciences (10%), NMP (8%) and Health Sciences (7%).
Our activity in European programs has been focused mostly on ICT and topics related to environment.

EDANYA group is a leading group concerning numerical modelling on geophysical flows and HPC techniques. EDANYA group is composed by 18 researchers, 7 of them with a permanent position at UMA, 2 non-permanent, 2 postdoc, 6 predoctoral students and 1 contracted personnel. In the last 5 years, they have published more than 40 papers in top rated international journals. In the same period, 3 PhD thesis have been defended. They have been participated and lidered 6 National research projects and in 4 EU projects: Modcompshock (ERC-ITN), ChEESE (H2020-INFRAEDI-2018), ARISTOTLE2-ENHSP, AGITHAR (COST action). Finally, they have leadered 19 transfer projects related with tsunami modelling.

Role of Partner in the Project

Tasks in the project: UMA, jointly with INGV and NGI, will provide a *"Probabilistic Tsunami Forecasting"* (PTF) workflow in WP6. UMA contributes to the project with a numerical code for tsunamis for seismically generated tsunamis, Tsunami-HySEA. Our role in the project is fundamental for a seamless integration of this code in the eFlows4HPC software stack. We will contribute in all tasks related with faster than real time numerical tsunami simulations, convenient and effective data storage, automatic and dynamic post-processing of the tsunami simulations, and probabilistic forecasting of the tsunami impact. We will also contribution in tasks in WP7 (exploitation) and WP8 (dissemination), we will collaborate when required by the leading institution in WP9 (management).

Previous relevant expertise: EDANYA-UMA, the research group on Differential Equations, Numerical Analysis and Applications (EDANYA) of the University of Málaga (UMA) has a wide experience on the numerical simulation and modelling of geophysical flows. EDANYA group has contributed in recent years to develop GPU (Graphics Processing Units) and multi-GPUs efficient implementations of several geophysical numerical tools related to tsunami modelling, making it possible to obtain accurate simulations quite faster-than-real-time computation time. In particular, Tsunami-HySEA model is currently used at INGV (Italy), IGN (Spain), JRC's European Crisis Management Laboratory and it is being used by SHOA (Chile), NOAA (USA), Seismic Network (Puerto Rico), IH Cantabria (Spain) and other international institutions. EDANYA group has also contributed to establish a theoretical framework for the design of efficient numerical schemes for non-conservative systems related to geophysical flows that has been used extensively by multidisciplinary researchers from different universities and institutions: Univ. of Trento (Italy), Univ. of Catania (Italy), Univ. of Seville (Spain), Univ. of Valencia (Spain), Univ. of Bordeaux (France), Univ. of Aachen (Germany), ETH (Switzerland), NOAA (USA), INGV (Italy), etc.
The EDANYA group is the leading group in Europe in the modeling and numerical simulation of tsunamis. The group is also a pioneer and worldwide reference in the implementation of this type of codes in GPU

and multi-GPU architectures. The in-house code for the simulation of the evolution and impact of tsunami waves, called Tsunami-HySEA, has been approved by the NTHMP (National Tsunami Hazard and Mitigation Program) to be used officially in risk assessment studies for tsunamis in USA, with funding from this program. The great computational efficiency of this code, together with its massively parallel implementation in multi-GPU architectures, has allowed it to integrate into Tsunami Early Warning Systems (TEWS) for the realization of simulations in extremely short times. At present the national systems of Spain and Italy implement Tsunami-HySEA as computing software. The work developed by the group EDANYA in the context of the numerical simulation of tsunamis and its implementation in GPU architectures, has led to its nomination as one of the 4 finalists of the NVIDIA Global Impact Award 2018.

Expected benefits: Our group has been working in Faster Than Real Time (FTRT) tsunami simulations for TEWS, Urgent Computing in Tsunami Science for Civil Proteccion applications and humanitarian aid, massively embarrassing tsunami computations for PTHA and, more recently, in the framework of ChEESE project, in HPC for Probabilistic Tsunami Forecasting (PTF). The expected benefits for UMA include greatly improved tsunami forecasting capabilities for early warning and emergency computing purposes using workflows for HPC, in particular within our core applications for PTF.

Qualification of key personnel

Jorge Macías Sánchez (male), holds a PhD in Mathematics (1998) from the Universidad de Málaga (Spain) and in Numerical Analysis (1998) from Université Paris VI (France). Full-time Associate Professor at UMA since 2001 and member of EDANYA Group since 1991. He is an expert researcher in the field of Numerical Analysis of Partial Differential Equations. His research interest mainly focussed in mathematical modelling and numerical simulation of geophysical flows, with particular focus on tsunami applications. More than 60 publications (orcid: 0000-0002-3010-8050), h-index=13 (ISI Web), h-index= 17 (Google Scholar). He has participated in more than 25 international and national research projects in the last 10 years and in more than 20 contracts since 2012, most of them related to tsunami research. Member of the Spanish "Technical Commission on Tsunami Risk". He is involved in the ICG/NEAMTWS since 2015 as member of the Spanish delegation and member of WG1 on tsunami modelling and also member of the WG2 at ICG/CARIBE-EWS. He has been a promoter of the Global Tsunami Model (GTM) scientific network since its early days, and of the AGITHAR COST Action (secondary proposer and member of the Management Committee).

Manuel J. Castro Díaz (male), PhD in Mathematics (1996) from the UMA. Full-time Associate Professor at UMA since 2000 and member of EDANYA Group since 1991. Specialist in the field of Numerical Analysis of PDEs. In particular, he focuses his research on the numerical analysis of non-conservative hyperbolic systems and the modeling of geophysical flows, with applications to the simulation of coastal currents, floods, avalanches, sediment transport, turbidity currents, generation and propagation of tsunamis, biphasic flows, magnetohydrodynamics, etc. He has also interest in the efficient implementation of the numerical models using multicore and multiGPU architectures and, more recently, on uncertainty quantification and data assimilation. He has been the IP of several Spanish research projects and international research contracts related to tsunami modelling. He also participated as researcher in several EU projects and MSCA-ITN network. He has published around 100 papers in JCR indexed journals and he has directed 7 PhD. Tesis at the universities of Málaga, Granada and Seville. Finally, he is member of the Scientific Computing Committee of ECCOMAS.

José Manuel González Vida (male), PhD in Applied Mathematics (2003) from the UMA. Full-time Associate Professor since 2008 in Applied Mathematics at the UMA. Member of EDANYA Group since 1997. His research is focused on the simulation of geophysical flows, under a multidisciplinary and applied framework. This framework led him to work with first level specialists in Geophysics, Marine Geology or Oceanography, to develop specific numerical schemes able to solve the numerous difficulties that are present in real technological problems and, more specially, in problems related to the numerical simulation of geophysical flows using finite volume schemes based on shallow water equations. Author of more than 40 research publications indexed in the WoS (h-index=11 ISI Web. WoS ResearcherID: E-2739-2012, ORCID 0000-0003-4996-1972). He has participated in 19 R&D projects funded through (European, National and Regional) competitive calls of public entities. PI of several research contracts (some of them co-leaded with Manuel J. Castro) with NOAA (U.S.A.), INGV (Italy) and JRC (European Union). Between

the results of these contracts it is remarkable the implementation of Tsunami-HySEA, a numerical model specifically created for Tsunami Early Warning Centers that has become the official tsunami numerical model of Italy and is being used by the Joint Research Center (European Union), NOAA in the United States and SHOA in Chile. This numerical model has been awarded with the important "Nvidia Global Impact Award" in March, 2018 as this model is being a game changer in the tsunami alert systems around the world and is one of the flagship models used in the H2020 Project ChEESE.

Publications/achievements relevant to the call

EDANYA-UMA since 1st October 2019 is providing, in collaboration with INGV, the Tsunami Service in ARISTOTLE-ENHSP. All Risk Integrated System TOwards Trans-hoListic Early-warning - European Natural Hazards Scientific Partnership) Consortium. The objective of this service is to provide 24*7 scientific advice on tsunamis to the Emergency Response Coordination Center (ERCC). Funding: DG-ECHO (European Commision).

Tsunami-HySEA code is currently been used by two National Tsunami Warning Centers (NTWC) in Europe (INGV, Italy and IGN, Spain). T-HySEA is also used a backup code at PMEL/NOAA (US).

EDANYA-UMA received the 2018 NVIDIA Global Impact Award in recognition for our pioneering work in GPU-based numerical model to accelerate tsunami simulations, producing Faster Than Real Time (FTRT) simulations in the framework of Tsunami Early Warning Systems (TEWS) with the aim of saving lives.

C. Escalante, M. Dumbser, M.J. Castro. "An efficient hyperbolic relaxation system for dispersive non-hydrostatic water waves and its solution with high order discontinuous Galerkin schemes". Journal of Computational Physics, Volume 394: 385-416, 2019.

J. Macías, M.J. Castro, S. Ortega, C. Escalante, J.M.González-Vida. "Performance benchmarking of Tsunami-HySEA model for NTHMP's inundation mapping activities". Pure and Applied Geophysics 174 (8), 3147-3183. 2017.

Relevant previous projects/activities

ChEESE (Centre of Excellence for Exascale in Solid Earth, 2018-2021) CoE in Exascale Computing (H2020), coordinated by BSC: The ChEESE proposal has 4 general objectives and a large list of specific objectives related to scientific, technical, and socio-economic challenges, among them to integrate around HPC and HDA transversal European institutions in charge of operational geophysical monitoring networks, Tier-0 supercomputing centers, academia, hardware developers, and third-parties from SMEs, Industry and public governance bodies (civil protection). In ChEESE, UMA is contributing to three pilot demonstrators related to tsunami hazard analysis and urgent tsunami computing, leading the pilot on Faster Than Real Time Tsunami Simulations and providing with the two tsunami flagship codes to the project.

ARISTOTLE-ENHSP (DG ECHO, 2018-2021; ECHO/SER/2018/683059): This project is the continuation of the ARISTOTLE multi-hazard early warning pilot project coordinated by INGV in Rome. In ARISTOTLE-ENHSP, UMA has implemented and is currently running the Tsunami Service, using as computational kernel the code Tsunami-HySEA.

MEGAFLOW: Relaxed non-hydrostatic multilayer models and high-order well-balanced numerical MEthods for GeophysicAl FLOWs. (Spanish MINECO project: 2019-2022): The main objective of this project is to introduce new multilayer non-hydrostatic mathematical models for the simulation of geophysical flows such as tsunami waves, storm surges, debris flows, landslides, etc. that can be used to model real-world problems. A relaxation strategy will be used to impose weakly the zero divergence constraint related to the non-hydrostatic pressure. New WB high order numerical methods to solve systems of balance laws or nonconservative hyperbolic systems will be developed and analyzed. Finally, HCP and Machine Learning techniques will be developed to improve the computational costs of the models. The overall objective is to develop geophysical flow solvers with improved physical quality able to produce FTRT simulations.

AGITHAR (Acceleration Global science In Tsunami HAzard and Risk Analysis). Ref.: OC-2018-1-22678. Funding: COST (European Cooperation in Science and Technology). EDANYA-UMA participates in this Action as Secondary Proposer and we are actively involved in several WGs. Among the aims of AGITHAR are determining gaps in scientific knowledge, methodological approaches and tools in order to achieve robust tsunami hazard and risk analysis, derive and agree on best practices and standards for probabilistic tsunami hazard and risk analysis or identifying issues and challenges to orient future tsunami modeling research.

Research contracts. Six contracts with the NOAA Center for Tsunami Research of the Pacific Marine Environmental Laboratory (NOAA), USA since 2012: Simulation for Landslide generated tsunamis; Develop a CUDA version of the tsunami NOAA model MOST. Contracts num.: WE133R12SE2276 2012-3, WE133R13SE1701 2013-14, WE133R15SE1487 2015-6, WE133R16SE1418 2016-7, WE133R17SE1310 2017-8, 133MJ19PNRMA0107 2019-20. Three contracts with the INGV (Italy) for the adaptation and setting up of the Tsunami-HySEA code for the INGV in 2013-14, 2015 and 2017. Other contracts with UNESCO, IGME (Spain), IGN (Spain), Seismic Network Puerto Rico (US) or INGV-Pisa (Italy).

Relevant infrastructures and networks

UMA has the following resources:

During 2020, UMA is going to host a new supercomputer with 8 homogeneous multi-GPU servers (8 NVIDIA Volta V100 with 32GB each one), 2x20 core CPUs and 512 MB RAM each one. Nodes are interconnected by Infiniband (100Mbps).

4.1.12. Istituto Nazionale di Geofisica e Vulcanologia.

Short name	INGV	
Type	Research Centre	
Website	www.ingv.it	
Country	Italy	

Description of the organization

INGV is the reference institution of the Italian Department of Civil Protection for the monitoring of seismic areas and active volcanoes and for tsunami warning. In 2013 INGV constituted the CAT (Centro Allerta Tsunami) that acts i) as Tsunami Service Provider in the NEAMTWS (since 2016) and ii) is operational at national level (since 2017). INGV leads the official programs for seismic, volcanic and tsunami hazard assessment promoted by the Italian government and actively participates in all European efforts on the same topics. INGV carries out fundamental research on the Solid Earth, including the physics of the Earth interior, earthquake generation processes, volcano dynamics, and on the Fluid Earth. INGV is the most significant Italian public research institution in the Earth Sciences and is the reference institution in these fields for the Italian Ministry of Education, Research and University and for several other government bodies

INGV has been involved in many EU granted projects and initiatives of which only the most relevant can be mentioned here. INGV hosts the headquarters of two relevant Research Infrastructures listed in the Roadmap of ESFRI: the Coordination Office (ECO) of the European Plate Observing System (EPOS) and the Legal Head Office of the European Multidisciplinary Seafloor Water Column Observatory (EMSO). Specifically concerning tsunamis, INGV coordinated the DG-ECHO co-funded project Probabilistic TSUnami Hazard MAPS for the NEAM Region (TSUMAPS-NEAM), which released the first homogeneous long-term Probabilistic Tsunami Hazard Assessment (PTHA) for earthquake-induced tsunamis for the coastlines of the NEAM region. INGV participated however to many other important European projects dealing with tsunamis (TRANSFER, MARSITE, STREST, ASTARTE). INGV has been co-coordinating both All Risk Integrated System TOwards Trans-boundary hoListic Early-warning (ARISTOTLE) and the following ARISTOTLE-ENSHP. INGV is one of the founding members of ORFEUS (Observatories & Research Facilities for European Seismology) and a main contributor of the EIDA European Integrated Data Archive (INGV is one of 11 nodes of the networks). It has been involved in several European projects concerning HPC computational seismology like SPICE (FP6), VERCE (FP7), EUDAT (FP7), EUDAT2020 (H2020), DARE (H2020), IMAGINE_IT (PRACE 9th).

Role of Partner in the Project

Tasks in the project: INGV will provide two workflows in WP6: jointly with NGI and Univ. Malaga, the "Probabilistic Tsunami Forecasting (PTF)" workflow and, jointly with BGS and ETH, the "Urgent Computing Integrated Services for EarthQuake (UCIS4EQ)" workflow. The PTF consists of seismic source ensemble initialization based on seismic data assimilation and earthquake parameter estimation, faster than real time numerical tsunami simulations, convenient and effective data storage, automatic and dynamic post-processing of the tsunami simulations, and probabilistic forecasting of the tsunami impact. Similarly, UCIS4EQ performs an ensemble of seismic wave simulations, providing probabilistic forecasting of the earthquake impact. In both cases, INGV main contribution will be for managing the initialization ensembles for uncertainty quantification and post-processing of simulation results to provide probabilistic forecasts

Previous relevant expertise: INGV is one of the major players worldwide in seismology and tsunami science. INGV is also in charge for tsunami warning and monitoring in the Mediterranean in the framework of the NEAMTWS, and for tsunami hazard mapping at the national level. INGV also co-coordinates the ARISTOTLE multi-hazard early warning service for ERCC. For further details, see the INGV organizational description, as well as the list of relevant projects. The experience from ChEESE on tsunami HPC and urgent computing is especially relevant.

Expected benefits: To provide realistic and unbiased intensity forecasts in the aftermath of an event is one of the activities of INGV in support of decision making. The expected benefits for INGV include greatly improved tsunami forecasting capabilities for early warning and emergency computing purposes using

workflows for HPC, in particular within our core applications for Probabilistic Tsunami Forecasting (PTF), as well as in the Urgent Computing Integrated Services for EarthQuake (UCIS4EQ) application. The new approaches to general workflows within natural hazard applications in eFlow4HPC (earthquakes and tsunamis) may demonstrate applicability to other project areas in the broader multi-hazard and multi-risk domain in the future, improving capabilities and catalyzing future innovation in neighboring applications not covered in this project, with the ultimate goal of disaster risk reduction.

Qualification of key personnel

Dr Jacopo Selva (male) was educated as a physicist; he has been working for more than 15 years in the field of the quantification of natural hazards from tsunamis, earthquakes and volcanic eruptions. He has published more than 50 journal papers (h-index 18 in Scopus, 20 in Google Scholar), participated in many International, European, and Italian scientific projects, also with coordination roles. He is member of the Steering Committee of CAT (Centro di Allerta Tsunami / Tsunami Warning Center) of INGV since 2013, and Responsible for tsunami hazard developments since 2017.

Alberto Michelini (male) PhD in geophysics, UC Berkeley, 1991. Research Director at INGV since 2003. Director of the National Earthquakes Center of INGV from 2013 to 2016. Author of more than 80 scientific publications in international journals. President of the ORFEUS ExeCom from 2008 to 2017, member of the EPOS Seismology Consortium Assembly. Expertise in many topics of seismology including real-time analysis and off-line analysis of large data volumes. Coordinator of the 2016-2018 ARISTOTLE EC project (ECHO/SER/2015/722144) and group leader of the 2018-2021 ARISTOTLE-ENHSP project (ECHO/SER/2018/683059).

Dr. Stefano Lorito (male) has a degree in physics from the University of Naples "Federico II" and a PhD in Geophysics from the University of Bologna "Alma Mater Studiorum". He has been working for more than 10 years now in the field of tsunami science: first dealing with inversion of tsunami and other geophysical data for imaging the tsunami source mainly for mega-thrust events; then with probabilistic tsunami hazard analysis and tsunami early warning. He is member of the governing board of the Italian national tsunami warning centre and also responsible for national tsunami hazard mapping. He has been involved in the ICG/NEAMTWS also as Vice-Chair 2017-19. He's been one of the main authors of the TSUMAPS-NEAM PTHA. He's been a promoter of the Global Tsunami Model (GTM) scientific network since its early days, and of the AGITHAR COST Action.

Dr. Manuela Volpe (female) is a researcher at INGV in Rome; her main research interests are focused on tsunami numerical modeling, analysis of the mechanisms of generation, propagation and inundation of tsunami waves, tsunami warning and probabilistic hazard assessment. She is involved in the development of the Italian tsunami warning center (CAT-INGV) and the construction of the Global Tsunami Model (GTM), as well as in National and International research projects.

Publications/achievements relevant to the call

Grezio, A., Babeyko, A., Baptista, M. A., Behrens, J., Costa, A., Davies, G., Geist, E.L., Glimsdal, S., González, F.I., Griffin, J., Harbitz, C.B., LeVeque, R.J., Lorito, S., Løvholt, F., Omira, R., Mueller, C., Paris, R., Parsons, T., Polet, J., Power, W., Selva, J., Sørensen, M.B., Thio, H. K. (2017). "Probabilistic Tsunami Hazard Analysis: Multiple sources and global applications". Reviews of Geophysics, 55. doi: 10.1002/2017RG000579

Lorito S, Selva J, Basili R, Romano F, Tiberti MM, Piatanesi A (2015), "Probabilistic Hazard for Seismically-Induced Tsunamis: Accuracy and Feasibility of Inundation Maps", Geophys. J. Int., 200 (1), 574-588, DOI:10.1093/gji/ggu408

Selva J, Tonini R, Molinari I, Tiberti MM, Romano F, Grezio A, Melini D, Piatanesi A, Basili R, Lorito S (2016), "Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA)", Geophys. J. Int., 205: 1780-1803, DOI: 10.1093/gji/ggw107

Michelini, A., Faenza, L., Lauciani, V., & Malagnini, L. (2008). "ShakeMap implementation in Italy". *Seismological Research Letters*, *79*(5), 688–697.

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinovic, D., Puglia, R., & Luzi, L. (2019). "The New

ShakeMap in Italy: Progress and Advances in the Last 10 Yr". Seismological Research Letters, 1–15. http://doi.org/10.1785/0220190130.

Relevant previous projects/activities

ChEESE (Centre of Excellence for Exascale in Solid Earth, 2018-2021) CoE in Exascale Computing (H2020), coordinated by BSC: The ChEESE proposal has 4 general objectives and a large list of specific objectives related to scientific, technical, and socio-economic challenges, among them to integrate around HPC and HDA transversal European institutions in charge of operational geophysical monitoring networks, Tier-0 supercomputing centers, academia, hardware developers, and third-parties from SMEs, Industry and public governance bodies (civil protection). In ChEESE, INGV are contributing to three pilot demonstrators related to tsunami hazard analysis and urgent tsunami computing, including the PTF.

TSUMAPS-NEAM (DG-ECHO, 2016-2018): The TSUMAPS-NEAM tsunami project was coordinated by INGV in Rome. The primary objective of the TSUMAPS-NEAM project was to develop homogenous regional tsunami hazard maps for the Mediterranean, the Black Sea, and the North East Atlantic region. TSUMAPS-NEAM is relevant for the present project, as parts of the software developed in TSUMAPS-NEAM formed the background for workflows used in the ChEESE CoE.

ASTARTE (EU FP7, 2014-2016): The ASTARTE tsunami project was coordinated by IPMA in Lisbon. In ASTARTE, INGV shared with NGI the responsibility of a work package for developing landslide and earthquake tsunami sources, and participated actively in several research activities focused on tsunami hazard and tsunami early warning methods.

ARISTOTLE (DG ECHO, 2016-2018; ECHO/SER/2015/722144): The ARISTOTLE multi-hazard early warning pilot project was coordinated by INGV in Rome. In ARISTOTLE, INGV lead the project in co-coordination with ZAMG and developed the Standard Operating Protocol, was responsible of the earthquake hazard activities, developed the IT platform for multi-hazard report preparation and participated to the operational activities for earthquakes, tsunami and volcano hazards.

ARISTOTLE-ENHSP (DG ECHO, 2018-2021; ECHO/SER/2018/683059): This project is the continuation of the ARISTOTLE multi-hazard early warning pilot project coordinated by INGV in Rome. In ARISTOTLE-ENHSP, INGV is group-leader and the project further elaborates the service provided in the previous project, adds the forest fires hazard and extends the coverage of the floods service to global level.

Relevant infrastructures and networks

INGV seismic monitoring and tsunami warning infrastructure will provide seismic solutions in real time feeding the proposed applications. The same infrastructures constitute in turn a test bed environment for the agreed use cases in this project.

Connection to EPOS will enhance sustainability of the services well beyond the project lifetime.

The project features a strong link to the activities by the project ARISTOTLE-ENHSP for both the generation of maps of strong ground shaking – crucial for impact assessment – and for the generation and propagation of tsunami.

4.1.13. Alfred Wegener Institute.

Short name	AWI	
Type	Research Centre	
Website	www.awi.de	
Country	Germany	

Description of the organization

The Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) is one of the world's leading polar research organisations and delivers significant contributions to the international research on climate, marine and coastal issues. AWI is a member of the Helmholtz Association of German Research Centres (HGF). It has an annual budget of more than 100 million Euros and a staff of more than 1000 employees. AWI keeps the German federal government updated on its research results and provides advice for the development of environmental policies. AWI conducts an extensive cooperation with numerous national and international institutions. Through memberships, AWI is embedded in a number of European Research initiatives, such as the European Polar Board (EPB), the European Marine Board (EMB), and the European Climate Research Alliance (ECRA). Furthermore, the office of the International Arctic Science Committee (IASC) and that of the International Coordination Office for Polar Prediction (ICO) are hosted by AWI.

The Climate Dynamics section of AWI is conducting research on understanding the functioning of the climate system, developing sea ice-ocean models based on unstructured mesh methods and their incorporation into climate models, analysis of climate predictions and projections and data assimilation.

Role of Partner in the Project

Tasks in the project: AWI will provide expertise on high resolution and ensemble climate model simulations, and provide its CMIP type climate model for the Pillar II case. AWI will lead initial collection of requirements for the end-to-end climate simulation/analysis workflow developed in Pillar II (WP5) and will coordinate evaluation of the Pillar II use cases. The state of the art OpenIFS-FESOM2 coupled climate model developed in AWI will be used for the Pillar II workflow. AWI will actively participate in the development of Pillar II use cases and integration of the results in the workflow repository and HPCWaaS interface.

Previous relevant expertise: AWI developed its climate model that participates in the CMIP6 activities. The ocean part of the climate model (FESOM2) is a next generation global ocean model that use unstructured meshes and it is developed from scratch by AWI scientist. AWI has its own supercomputing facilities and has extensive experience in running ocean and climate model simulations/post processing tasks on different HPC systems. The set of tools (ESM-Tools) developed in AWI to unify model infrastructure, giving a common framework for downloading, compiling, running and organizing coupled or standalone models, so there is an experience in formalisation of parts of the climate model run workflow.

Expected benefits: The expected benefits for AWI are in increased capabilities to gain robust scientific results in the area of climate modelling using less computational resources. We also expect to overcome some of the challenges imposed by the increased amount of data associated with higher spatial resolution of Earth System Models, by using dynamically adjusted workflows and ML technologies. Implementation of the OpenIFS/FESOM2 as part of the workflow repository and HPCWaaS will increase the adoption rate of the AWI coupled model and open new opportunities for cooperation and improvement of the model. The project will also help to identify performance bottlenecks and prepare OpenIFS/FESOM2 for exascale era.

Qualification of key personnel

Prof. Dr. Thomas Jung (male), PhD in Atmospheric Physics, Institute for Marine Research, Kiel. Head of Climate Dynamics Section at AWI, where he is leading the climate analysis, modelling and prediction activities. He is also a full professor for physics of the climate system at the University of Bremen. Before joining the AWI, Prof Jung worked for 10 years at the European Centre for Medium Range Weather Forecasts (ECMWF). He is a Member of the Scientific Advisory Committee of ECMWF, Member of the

Scientific Advisory Board of the John von Neumann-Institute for Computing and Chair of the Scientific Advisory Committee of the German Weather Service. Prof Jung coordinate project Advanced Earth System Modelling Capacity (ESM) funded by the Helmholtz Association. He is spokesperson of AWI's research programme. He acts as the chair of the Polar Prediction Project of the World Meteorological Organization's (WMO) World Weather Research Programme. Prof Jung is also a member of WMO's EC Panel of Experts on Polar Observations, Research and Services (EC-PORS), which oversees the implementation of the Global Integrated Polar Prediction System (GIPPS). He is the (co)-author of more than 120 publications in the peer-reviewed literature.

Prof. Dr. Sergey Danilov (male), PhD in Physics and Mathematics, Acoustical Institute of Russian Academy of Sciences. Principal research scientist of Climate Dynamics Section at AWI. He is also professor of Numerical Modeling in the Jacobs University. Prof. Danilov is leading the development of Finite-Volume Sea ice – Ocean circulation model (FESOM2) and its sea-ice component FESIM, the first multi-resolution models worldwide that are practically used for climate research. His major expertise lies in numerical methods on unstructured meshes, applied mathematics and geophysical turbulence, with particular focus on mathematics and physics related to modelling high-resolution dynamics. His recent work involves questions of stability and convergence of sea-ice dynamics solvers, which present a challenge at high resolution.

Dr. Dmitry Sidorenko, PhD in Physical Oceanography. Senior scientist of Climate Dynamics section at AWI. Dr. Sidorenko is a main developer of Finite-Volume Sea ice – Ocean circulation model (FESOM2). He is also the main developer of several versions of AWI climate model. He has more than 15 years of experience running ocean and climate simulations on different HPC systems. His main research interests include numerical algorithms on unstructured meshes and multiresolution climate simulations.

Dr. Dirk Barbi, PhD in Physics. Senior scientist of Climate Dynamics section at AWI. Dr. Barbi is one of the developers of the AWI climate model. He has more than 15 years of experience of developing and supporting different parallel geophysical codes on HPC systems. Currently he is leading the development of ESM-Tools, a python based tools to abstract some basic parts of climate modelling workflow.

Publications/achievements relevant to the call

Danilov, S., Sidorenko, D., Wang, Q., & Jung, T. (2017). "The Finite-volumE Sea ice-Ocean Model (FESOM2)". Geoscientific Model Development, 10(2), 765.

Sidorenko, D., Goessling, H. F., Koldunov, N. V., Scholz, P., Danilov, S., Barbi, D., et al (2019). "Evaluation of FESOM2.0 coupled to ECHAM6.3: Pre-industrial and HighResMIP simulations". Journal of Advances in Modeling Earth Systems, 11. doi.org/10.1029/2019MS001696

Koldunov, N. V., Aizinger, V., Rakowsky, N., Scholz, P., Sidorenko, D., Danilov, S., and Jung, T.: "Scalability and some optimization of the Finite-volumE Sea ice–Ocean Model, Version 2.0 (FESOM2)", Geosci. Model Dev., 12, 3991–4012, doi.org/10.5194/gmd-12-3991-2019, 2019.

Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., Sein, D., Smolentseva, M., Rakowsky, N., and Jung, T.: "Assessment of the Finite-volumE Sea ice-Ocean Model (FESOM2.0) – Part 1: Description of selected key model elements and comparison to its predecessor version", Geosci. Model Dev., 12, 4875–4899, doi.org/10.5194/gmd-12-4875-2019, 2019.

Rackow, T., Sein, D. V., Semmler, T., Danilov, S., Koldunov, N. V., Sidorenko, D., Wang, Q., and Jung, T.: "Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM1.0", Geosci. Model Dev., 12, 2635-2656, doi.org/10.5194/gmd-12-2635-2019, 2019.

Relevant previous projects/activities

CMIP6 [WCRP] Coupled Model Intercomparison Project. CMIP6 coordinates somewhat independent model intercomparison activities and their experiments which have adopted a common infrastructure for collecting, organizing, and distributing output from models performing common sets of experiments. This information has proved invaluable in preparing high profile reports assessing our understanding of climate and climate change (e.g., the IPCC Assessment Reports). AWI provide climate model simulations for CMIP6, including high resolution ones performed in accordance with HighResMIP protocol.

ESM [Helmholtz Association, 2017-2020**]** Advanced Earth System Modelling Capacity. The project comprises eight Helmholtz Research Centers and aims to improve the representation of the components of the Earth system and their coupling, as well as to perform a series of selected numerical experiments to address Grand Challenges (Frontier Simulations). A long-term strategy for the development of an Earth System Modelling capacity is also an objective of the project. Thomas Jung (AWI) is the coordinator of the project.

PL-EESM [Helmholtz Association, 2020-2022] The Pilot Lab Exascale Earth System Modelling. Explores specific concepts to enable exascale readiness of Earth System models and associated workflows in Earth System science. PL-EESM provides a new platform for scientists of the Helmholtz Association to develop scientific and technological concepts for future generation Earth System models and data analysis systems.

PRIMAVERA [EU H2020, 2015-2019] PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment). The goal of PRIMAVERA is to deliver novel, advanced and well-evaluated high-resolution global climate models (GCMs), capable of simulating and projecting regional climate with unprecedented fidelity, out to 2050. The Climate Dynamics section provide high resolution and frontier resolution simulations.

TRR181 [DFG 2016-2020] Energy Transfers in Atmosphere and Ocean. As part of the subproject S1 (Diagnosis and Metrics in Climate Models) Climate Dynamics section of AWI develop ocean model diagnostics for CMIP6 models as well as specific diagnostics for models on unstructured meshes, that can be used for scalable parallel data analysis.

Relevant infrastructures and networks

The AWI is an associate of the German Climate Computing Centre (DKRZ). Therefore, AWI scientists have dedicated access to the supercomputing and data storage facilities at DKRZ. Furthermore, the scientific computing department of AWI provides high-performance computing facilities for model development and testing purpose (CRAY CS 400 with 11.232 cores). Through ESM project Climate Dynamics section in AWI can apply for resources on dedicated compute partition on the Tier-0/1 system (JUWELS) Jülich Supercomputing Centre (JSC).

4.1.14. Eidgenoessische Technische Hochschule Zuerich

Short name	ETH Zürich
Type	University
Website	www.ethz.ch
Country	Switzerland

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Description of the organization

Freedom and individual responsibility, entrepreneurial spirit and open-mindedness: ETH Zurich stands on a bedrock of true Swiss values. Our university for science and technology dates back to the year 1855, when the founders of modern-day Switzerland created it as a centre of innovation and knowledge.
At ETH Zurich, students discover an ideal environment for independent thinking, researchers a climate which inspires top performance. Situated in the heart of Europe, yet forging connections all over the world, ETH Zurich is pioneering effective solutions to the global challenges of today and tomorrow.
Facts and figures:

- 21,400 students, including 4,180 doctoral students, from over 120 countries
- 530 professors
- 407 spin-offs since 1996
- 21 Nobel Prize winners, 2 Fields Medal winners, 2 Pritzker Prize winners
- 109 patent applications and 205 invention reports every year
- 11thin THE ranking (2018/19), 7thin QS ranking (2018/19), 19thin ARWU ranking (2018/19).

Role of Partner in the Project

Tasks in the project: Mainly responsible for Task 6.3 (Development of the Pillar III: earthquakes). The focus of this task is twofold: will focus on (1) the development of novel features for UCIS4EQ as well as features that can be added as modules to it and (2) generation of a database of local 3D velocity models.

Previous relevant expertise: The Seismology & Wave Physics Group at ETH Zurich has ample expertise in the development and use of HPC numerical wave propagation codes, notably SES3D and Salvus. These have been used primarily for the purpose of full seismic waveform inversion from local to global scales, which is also the central theme of our involvement in this project.
Furthermore, the group has played a leading role in the development of novel techniques of uncertainty quantification in full seismic waveform inversion. The most recent developments include a nonlinear nullspace shuttle approach for the computation of alternative plausibe Earth models.

Expected benefits: The expected benefits for the project are twofold: (1) A suite of Earth models for various regions of interest, in the context of seismic hazard, and (2) a forward modelling machinery that allows us to produce seismic ground motion models at high frequencies.

Qualification of key personnel

Prof. Andreas Fichtner (male) is leading the Seismology & Wave Physics Group at ETH Zurich. He is an expert in inverse theory, numerical wave propagation, seismic interferometry, and high-performance computing. He is the author of 4 books and of more than 70 articles in international journals.

Dr. Marta Pienkowska-Côte (female) is a postdoc in the Seismology & Wave Physics Group at ETH Zurich. She is an expert in numerical wave propagation, specifically in the context of ground motion prediction and the effects of 3D structure and topography.

Dr. Christian Boehm (male) is a senior scientist in the Seismology & Wave Physics Group at ETH Zurich. He is one of the main developers of the Salvus spectral-element solver, which will be the workhorse of the high-performance numerical wave propagation simulations.

Publications/achievements relevant to the call

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D. A., Knepley, M. G.,

Fichtner, A., 2019. "Modular and flexible spectral-element waveform modelling in two and three dimensions". Geophysical Journal International 216, 1675-1692, doi: 10.1093/gji/ggy469.

Fichtner, A., Zunino, A., 2019. "Hamiltonian Nullspace Shuttles". Geophysical Research Letters, doi: 10.1029/2018GL08931.

Fichtner, A., van Herwaarden, D.-P., Afanasiev, M., Simute, S., Krischer, L., Cubuk-Sabuncu, Y., Taymaz, T., Colli, L., Saygin, E., Villasenor, A., Trampert, J., Cupillard, P., Bunge, H.-P., Igel, H., 2018. "The Collaborative Seismic Earth Model: Generation I". Geophysical Research Letters 45, doi: 10.1029/2018GL077338.

Fichtner, A., 2011. "Full seismic waveform modelling and inversion". 350pp, Springer-Verlag, Heidelberg.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., Villasenor, A., 2013. "Multi-scale full waveform inversion". Geophysical Journal International 194, 534-556.

Relevant previous projects/activities

ERC Starting Grant: The Collaborative Seismic Earth Model (Development of global multi-scale Earth Model).

Participation in EU Centre of Excellence ChEESE: Leading work package on urgent seismic computing.

Grant from the Swiss National Supercomputing Centre: Salvus (Development of massively parallel seismic wave propagation code).

Relevant infrastructures and networks

Does not apply.

4.1.15. Siemens AG

Short name	Siemens	
Type	Industry	
Website	www.siemens.com	
Country	Germany	

Description of the organization

Siemens AG (headquartered in Berlin and Munich) is a global powerhouse in electronics and electrical engineering. Operating in the fields of automation, electrification and mainly also digitalization, Siemens holds leading market positions in all its business areas. The company has roughly 379,000 employees – of which 41,000 were newly hired in FY18 and of which 117,000 or 31% (newly 4,700) are based in Germany – working to develop and manufacture products, design and install complex systems and projects, and tailor a wide range of solutions for individual requirements. For more than 171 years, Siemens stands for technological excellence, innovation, quality, reliability and internationality. In FY18, Siemens had revenue with business in more than 200 countries of €83.04 billion. In FY17, arising from Siemens software and digital services alone, €5.2 billion could be achieved (e.g. via MindSphere), making it a growth rate in this area of 20%.

Innovation is Siemens' most important growth and productivity driver. In FY18, the company invested €5.6 billion – 6.7 percent of its total revenue – in research and development (R&D) to stay at the forefront of technological progress. With plans for further raising these spendings in FY19, this means a growth rate of spending of 40%+ compared to FY14. Today the company employs some 43.400 researchers and developers as well as 10.900 apprentices and students in dual study programs, who work on innovations that secure existing business and open up new markets. Collaborations are an indispensable mean of developing strategically important technologies. By discussing, sharing, and implementing ideas with scientists from outside the company, especially with 25 preferred universities and numerous research centers, Siemens researchers keep abreast of the latest findings resulting from fundamental and applied research all over the world.

Siemens' patent portfolio in 2018 consists of 63,800 patents worldwide, with 3,850 patent applications and 7,300 inventions. According to statistics for patent applications from 2015, Siemens was number five in Europe, ranked number six in Germany and 33 in the U.S.
With over 8,100 employees worldwide (FY17), Corporate Technology (CT) plays a key role in R&D at Siemens. In research centers located in many different countries, CT works closely with the R&D teams in the Siemens´ Divisions. Corporate Technology is networked to facilitate efficient collaboration between its various sites around the world and with the rest of the Company. Its principal research operations are in close proximity to worldwide business activities and technology hotspots: Germany, the U.S., Austria, Russia, India, China and UK are among the most important sites. Establishing own research centers is combined with an intensive cooperation with top universities – in total 25 preferred university partners. This allows responding quickly and precisely to the demands of the customers and to provide solutions that meet local needs.

The CT organization provides expertise regarding strategically important areas to ensure the company's technological future, and to acquire patent rights that safeguard the company's business operations. Against the background of megatrends such as climate change, urbanization, globalization, and demographic change, CT focuses on innovations that have the potential to change the rules of the game over the long term in business areas that are of interest to Siemens. CT covers a wide range of technology fields including the subject areas materials and microsystems, production methods, security, software and engineering, power engineering, sensors, automation, medical information systems and imaging methods, information and communication technologies, the extraction and processing of raw materials, and off-grid power generation. So called SMART ("Simple, Maintenance-friendly, Affordable, Reliable and Timely to market") solutions implement new technologies in a manner that renders them competitive in low-price markets.

Role of Partner in the Project

Tasks in the project: WP4: Provide industrial use case and transfer results to industrial application.

Previous relevant expertise: Siemens AG Corporate Technology is working on model order reduction as a key technology to enable Digital Twins. Currently we develop simulation based solutions to support the operation of Siemens products like power transformers, infrastructure networks (energy/ water networks) or railway infrastructure.

Expected benefits: Siemens AG will benefit from eFlows4HPC in several ways. First the (offline) creation of reduced order models is currently a cumbrous process and only realizable in the industrial environment for a very limited application field (mainly pure thermal problems). However there is a high demand for operation support (e.g. decision support system) for more complex systems (e.g. oil cooled power transformers). The methods and workflows developed in eFlows4HPC promise to make these kind of industrial problems accessible.

In another way the availability of small and fast simulation models can boost the development of intelligent automation components, as more and more automation components are equipped with substantial computing power. Small and fast ROM-models can thus be executed on such edge devices to foster fast model predictive control.

In industrial environments the access to sensor data is only limited. Therefore fast physics based simulations can be used to enhance the quality of the sensor signal and give a proper estimation of uncertainties, by simulating many different alternatives – a task optimally suited for massive parallel computation.

Qualification of key personnel

Dr. Stefan Boschert (male) graduated in physics from the University of Freiburg i.Br. After his PhD in 2000 he worked for an engineering company in Basel, Switzerland. Since 2004 he is working at Siemens CT currently in the technology field Simulation & Digital Twin. As senior key expert he is responsible for simulation-based system development and virtual design, which includes multiphysics simulation, process models and integrated data & simulation model-based approaches. He has expertise in the set-up and implementation of mechatronic design processes and its extension to cyberphysical systems. Currently he focuses his research on developping simulation solutions supporting the operation of physical assets, thus combining physics-based simulation methods with data driven approaches. Further, he was manager of the department's contributions to the German government (BMBF) funded project ABBILD and mecPro² and the EU funded projects EXCITING and TERRIFIC. He will act as responsible for the Siemens team.

Dr. Arianna Bosco (female) studied Mathematical Engineering at the Università di Roma Tor Vergata (Bachelor) and at the Politecnico di Torino (Master). She has a PhD in Mechanical Engineering from the RWTH University in Aachen. For more than 12 years she has been working in the field of Computational Fluid Dynamics (CFD) with experience in all flows regimes from subsonic to hypersonic. She worked for more than 5 years at MTU Aero Engines in Munich in the Aerodynamic department. Her work focused on validation of new methods especially for aeroelasticity with a focus on Flutter and Forced Response phenomena. Since 2017 she works in R&D in Siemens Corporate Technology in the fields of CFD, structural mechanics and Model Order Reduction for both simulation fields. The aim of the Model Order Reduction research is to build digital twins of Siemens assets which runs parallel to their operation and can enhance predictive maintenance and optimize the performance.

Publications/achievements relevant to the call

Boschert S., Rosen R. (2016) "Digital Twin—The Simulation Aspect". In: Hehenberger P., Bradley D. (eds) Mechatronic Futures. Springer.

Rosen R. Boschert S., Fischer J. (2019): "Next Generation Digital Twin: an Ecosystem for Mechatronic Systems?" In: Proc. 8th IFAC Symp. On mechatronic systems, Vienna.

Hartmann D., Herz M., Wever U.: (2018) "Model order reduction a key technology for digital twins": In Keiper W. Milde A.,Volkwein S. (eds): Reduced-Order Modeling (ROM) for Simulation and

Optimization, Springer.
Relevant previous projects/activities
ALICE III (national BMBF, 08/18-07/21): developing ML methods that solve complex industrial problems, efficient training of robust ML methods for multi-component industrial plants. SensformerAdvanced : New product line to add connectivity and enhanced operational control to power transformers (www.siemens.com/sensformer).
Relevant infrastructures and networks
None.

4.1.16. Stiftelsen Norges Geotekniske Institutt

Short name	NGI	
Type	Research Centre	
Website	www.ngi.no	
Country	Norway	

Description of the organization

The Norwegian Geotechnical Institute (NGI) is a leading international centre for research and consulting within the geosciences, and a private not-for-profit national research foundation. It constitutes an excellent environment for establishing a group of both high individual and broad interdisciplinary quality. NGI's core mission is to aid development within research, consultancy, and contributions to education in the field of engineering-related geosciences. In 2003, NGI was awarded the national research Center of Excellence (CoE) status, the International Center for Geohazards (ICG). NGI is in the international forefront in the field of tsunami modelling within Pillar III of the proposal. NGI has participated in tsunami research over the last decade through major EU programs (TRANSFER, ASTARTE, TSUMAPS, ITN-SLATE, and ChEESE). NGI has also been heavily involved in regional and global tsunami assessments assisting major stakeholders such as the United Nations International Strategy for Disaster Reduction (UN-ISDR) and the World Bank's Global Facility of Disaster Risk Reduction (GFDRR). Furthermore, NGI hosts one of the world's largest geotechnical laboratories, and was recently awarded a new national set of geotechnical benchmark test sites.

Role of Partner in the Project

Tasks in the project: NGI will, jointly with INGV and Univ. Malaga, provide a Probabilistic Tsunami Forecasting workflow in WP6. This work will consist of seismic source ensemble initialization based on seismic data assimilation and earthquake parameter estimation, Faster Than Real Time (FTRT) numerical tsunami simulations, Probabilistic Tsunami Forecasting (PTF), convenient and efficient data storage, and automatic and dynamic post-processing of the tsunami simulations. In particular, NGI will take responsibility for revising and adopting existing PTF workflows, interfacing the HPC centers for utilizing general workflow managers such as PyCOMPSs. Moreover, we will develop flexible post processing routines for analyzing post event data, providing event diagnostics, as well as other deep post processing tools like, for instance, scenario disaggregation local probability curves.

Previous relevant expertise: For more than 20 years, NGI has been a major player within the field of numerical tsunami modelling. For further details, see the NGI organizational description, as well as the list of relevant projects. The experience from ChEESE on tsunami HPC and urgent computing is especially relevant.

Expected benefits: The expected benefits for NGI include greatly improved modelling capabilities using workflows for HPC and early warning, in particular within our core applications on tsunami simulations. The new approaches to general workflows within natural hazard applications in eFlow4HPC (earthquakes and tsunamis) may demonstrate applicability to other project areas covered by NGI in the future, improving capabilities and catalyzing future innovation in neighboring applications not covered in this project.

Qualifications of key personnel

Dr Finn Løvholt (male) is a tsunami specialist with a background in fluid mechanics. He received his MSc (1998) in Physics at the Norwegian University of Life Sciences, and PhD (2008) in Mechanics at the University of Oslo. His main expertise comprises numerical modelling of the generation, propagation, and run-up of tsunamis, including tsunami hazard and risk assessment. A strong emphasis in his research has been on landslide tsunami generation and wave mechanics, as well as tsunami hazard assessment. Løvholt has wide experience in tsunami research as well as being the PI in multiple large tsunami research modelling projects; he is one of the proposers behind the Global Tsunami Model, and has for a number of years coordinated the global tsunami hazard mapping for the UN-DDR. He is also a member of the UN Global Risk Assessment Framework (GRAF) expert group. He is NGI's PI in the ChEESE CoE, where he

is leading the pilot demonstrator on probabilistic tsunami hazard, and contributing on the tsunami urgent computing and probabilistic tsunami forecasting. Moreover, he has participated in three former EU tsunami projects (ASTARTE, TSUMAPS, and TRANSFER), and is presently one of the PI's in the ongoing ITN-SLATE project on landslides. He is presently the supervisor for two PhD students, and has previously co-supervised 3 PhD students and 4 postdocs. He has co-authored 54 papers, 23 as first author, in peer reviewed journals.

Dr. Sylfest Glimsdal (male) received his PhD (2007) and MSC (2001) in Fluid Mechanics from University of Oslo. He has a background from research and advisory projects in fluid dynamics. His key qualifications comprise wave mechanics, storm surge, tsunami generation, propagation and inundation, numerical analysis and simulation. A strong emphasis in his research has been on model verification, parametric studies and model development as well as local and global tsunami hazard assessment from earthquake sources. He is participating in the ChEESE CoE, contributing to the pilot demonstrator on probabilistic tsunami hazard, as well as on urgent tsunami computing and probabilistic tsunami forecasting. He has co-authored more than 28 papers, 8 as lead author, in peer-reviewed journals.

Dr. Steven J. Gibbons (male) is a geoscientist with a background in fluid mechanics and many years of experience with computational workflows for data processing and real-time seismic monitoring. He received his PhD in fluid mechanics from the University of Leeds in 1999 and developed large-scale codes for numerical simulations of planetary fluid flow and magnetic field generation (1998-2002). Between 2002 and 2019, he worked in seismic monitoring for Nuclear-Test-Ban Treaty verification and developed algorithms and workflows for signal detection, parameter estimation, and pattern recognition. Since June 2019 he has worked on numerical simulations of tsunamis and landslides for hazard analysis, and machine learning. He has been Principal Investigator on numerous projects within signal processing, data analysis, and probabilistic models for customers in Europe and the United States, and has participated in EU projects ARISE, EPOS, and ChEESE. He has co-authored 48 papers in peer-reviewed journals, 26 as first author.

Publications/achievements relevant to the call

Løvholt, F., Lorito, S., Macías, J., Volpe, M., Selva, J., Gibbons, S. (2019), "Urgent tsunami computing, in International Conference for High Performance Computing", Networking, Storage and Analysis 2019, IEEE/ACM HPC for Urgent Decision Making (UrgentHPC), doi: 10.1109/UrgentHPC49580.2019.00011

Gibbons, S.J., Kværna, T., Harris, D.B., Dodge D.A. (2016); "Iterative Strategies for Aftershock Classification in Automatic Seismic Processing Pipelines". Seismological Research Letters; 87 (4): 919–929. doi: 10.1785/0220160047

Grezio, A., Babeyko, A., Baptista, M. A., Behrens, J., Costa, A., Davies, G., Geist, E.L., Glimsdal, S., González, F.I., Griffin, J., Harbitz, C.B., LeVeque, R.J., Lorito, S., Løvholt, F., Omira, R., Mueller, C., Paris, R., Parsons, T., Polet, J., Power, W., Selva, J., Sørensen, M.B., Thio, H. K. (2017). "Probabilistic Tsunami Hazard Analysis: Multiple sources and global applications". Reviews of Geophysics, 55. doi: 10.1002/2017RG000579

Davies, G., Griffin, J., Løvholt, F., Glimsdal, S., Harbitz C., Thio, H.K., Lorito, S., Basili, R., Selva, J., Geist, E., and Baptista, M.A. (2018), A" global probabilistic tsunami hazard assessment from earthquake sources, Geological Society", London, Special Publications, 456, doi:10.1144/SP456.5

Gylfadottir, S.S., Kim, J., Helgason, J.K., Brynjolfsson, S., Hoskuldsson, A., Johannesson, T., Harbitz, C.B., Løvholt, F. (2017) "The 2014 Lake Askja rockslide tsunami – optimization of landslide parameters comparing numerical simulations with observed run-up", J Geophys. Res. Oceans, 122, doi:10.1002/2016JC012496.

Relevant previous projects/activities

ChEESE (Centre of Excellence for Exascale in Solid Earth, 2018-2021) CoE in Exascale Computing (H2020), coordinated by BSC: The ChEESE proposal has 4 general objectives and a large list of specific objectives related to scientific, technical, and socio-economic challenges, among them to integrate around HPC and HDA transversal European institutions in charge of operational geophysical monitoring networks, Tier-0 supercomputing centers, academia, hardware developers, and third-parties from SMEs, Industry and public governance bodies (civil protection). In ChEESE, NGI are contributing to three pilot demonstrators related to tsunami hazard analysis and urgent tsunami computing.

SLATE (ITN-MSC, H2020, 2017-2021, coordinated by Univ. Bremen): A key objective in SLATE is to understand key factors in triggering submarine landslides, the subsequent motion and evolution of failed material, as well as ensuing geohazards, e.g. tsunamis. The objectives of the tsunami activities in SLATE, led by NGI, are (i) on fundamental of landslides tsunamis through case studies and (ii) development of new coupled landslide tsunami models.

TSUMAPS (DG-ECHO, 2016-2018): The TSUMAPS tsunami project was coordinated by INGV in Rome. The primary objective of the TSUMAPS project was to develop homogenous regional tsunami hazard maps for the Mediterranean, the Black Sea, and the North East Atlantic region. TSUMAPS is relevant for the present project, as parts of the software developed in TSUMAPS formed the background for workflows used in the ChEESE CoE.

ASTARTE (EU FP7, 2014-2016): The ASTARTE tsunami project was coordinated by IPMA in Lisbon. In ASTARTE, NGI shared had the responsibility of a work package for developing landslide and earthquake sources.

Tsunamis induced by large landslides (TsunamiLand, Research Council of Norway 2014 - 2017): One out of seven national talented young researcher project in natural sciences granted by the Research Council of Norway (RCN) with 7% success rate. The project was granted to NGI's PI in this project (Finn Løvholt). In TsunamiLand we have developed new models for landslide run-out (BingClaw) and coupling them to tsunami generation, hence improving both numerical models for tsunami generation and the understanding for how they happen.

Relevant infrastructures and networks

None.

4.2. Third parties involved in the project (including use of third party resources)

Participant 1: BSC

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	Y
The High Performance Computing research group of the Computer Architecture Department at the Universitat Politècnica de Catalunya (UPC) is the leading research group in Europe in topics related to high performance processor architectures, runtime support for parallel programming models and performance tuning applications for supercomputing. There is a signed Collaboration Agreement between the UPC and the BSC establishing the framework of the relationship between these two entities. According to this agreement, several researchers of the UPC are made available to the BSC to work on projects. Dr. Yolanda Becerra is an associate professor at the UPC. She carries out her research activities in association with the Barcelona Supercomputing Center - Centro Nacional de Computación (BSC) on the BSC premises.	
Does the participant envisage that part of the work is performed by International Partner (Article 14a of the General Model Grant Agreement)?	N

Participant 2: CIMNE
No third parties involved, no subcontracts introduced.
Participant 3: FZJ
No third parties involved, no subcontracts introduced.
Participant 4: UPV
No third parties involved, no subcontracts introduced.
Participant 5: ATOS
No third parties involved, no subcontracts introduced.
Participant 6: DtoK Lab
No third parties involved, no subcontracts introduced.
Participant 7: CMCC
No third parties involved, no subcontracts introduced.
Participant 8: INRIA

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties	Y
In this Project, INRIA represents the Team Cardamom. Cardamom Team is a Joint Research Unit for which INRIA will represent Bordeaux INP, a Third Party linked to INRIA in future Grant Agreement and Consortium Agreement: Bordeaux INP as a Third Party linked to INRIA, will carry out part of the work attributed by the future Grant Agreement. Dr Beaugendre and Dr Barral (employed by Bordeaux INP) will work on WP4. The members of the third party will provide contributions to task 4.4 and in particular error estimation and adaptation methods required in the construction of efficient realistic models for the uses cases defined in Task 4.3.	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

Does the participant envisage that part of the work is performed by International Partner (Article 14a of the General Model Grant Agreement)?	N

Participant 9: SISSA
No third parties involved, no subcontracts introduced.
Participant 10: PSNC
No third parties involved, no subcontracts introduced.
Participant 11: UMA
No third parties involved, no subcontracts introduced.
Participant 12: INGV
No third parties involved, no subcontracts introduced.
Participant 13: AWI
No third parties involved, no subcontracts introduced.
Participant 14: ETHZ
No third parties involved, no subcontracts introduced.
Participant 15: Siemens
No third parties involved, no subcontracts introduced.
Participant 16: NGI
No third parties involved, no subcontracts introduced.

Section 5: Ethics and Security

5.1 Ethics

The eFlows4HPC partners will comply with the respective national laws and regulations concerning protection of personal data, which are in line with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data (as well as Regulation (EU) 2016/679 and Directive (EU) 2016/680).

5.2 Security

The eFlows4HPC project will not involve any security issues. Concretely:

- NO activities or results raising security issues
- NO 'EU-classified information' as background or results.

Annex A: Letters of Support

We attach the letters of support of 4 different entities that are interested in the results of the eFlows4HPC project:

1. **Center of Excellence for Exascale in Solid Earth**, letter from Arnau Folch, its coordinator. The CoE has large interest in the eFlows4HPC project results since two of the ChEESE flagship codes will be used in the workflows from pillar III and the results obtained in eFlows4HPC would have large benefits for ChEESE services.

2. **Center of Excellence EXCELLERAT**, letter from Bastian Koller, its coordinator. The CoE is highly interested in the results that could be obtained by pillar I, which deals with the integration of state-of-the-art solvers with the eFlows4HPC software stack in the area of manufacturing. The EXCELLERAT CoE has interest in the results of this pillar as a new type of workflow of interest for their community.

3. **Center of Excellence ESiWACE**, letter from Joachim Biercamp, its coordinator. Given that eFlows4HPC devotes Pillar II to climate workflows, ESiWACE has large interest in learning about the project developments and on joint collaborations between the two initiatives.

4. **Argonne National Laboratory**, letter from Ian Foster, a recognized expert in the area of distributed computing, who is currently applying machine learning to workflows developed with tools similar to the proposed in eFlows4HPC (Swift and Parsl) to problems with large climate and weather datasets. Professor Foster states in his letter his interest for the project and willingness to collaboration in case the proposal is funded.

Barcelona, 11th December 2019

To: Rosa M Badia
Coordinator of the eFflows4HPC proposal
Barcelona Supercomputing Center (BSC)

Subject: Letter of Support to eFflows4HPC

Dear Dr. Badia,

As coordinator of the **Center of Excellence for Exascale in Solid Earth** (ChEESE), I herewith express our strong interest and motivation to collaborate with the researchers involved in the proposal entitled "*Enabling dynamic and Intelligent workflows in the future EuroHPC ecosystem - eFflows4HPC*", submitted to call EuroHPC-02-2019. The main objective of ChEESE is the preparation of 10 Community flagship European codes in the Solid Earth domain for the upcoming pre-Exascale (2020) and Exascale (2022) supercomputers in order to enable services oriented to society on critical aspects of geohazards like hazard assessment, urgent computing, and early warning forecast.

eFlows4HPC intends to deliver a workflow software stack and additional set of services that will enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The project will integrate existing workflow interfaces, programming models, machine learning and data analytic libraries to provide a uniform platform that enable the exploitation of future large-scale systems. The project will demonstrate through three application pillars with high industrial and social relevance: manufacturing, climate and urgent computing for natural hazards, how the realization of forthcoming efficient HPC and data centric applications can be developed with new workflow technologies. The project aims also at delivering methodologies and tools that widens and makes easier the use of workflows by existing and new HPC communities and users.

Two of the ChEESE flagship codes (THySEA and Salvus for tsunami and earthquake modelling respectively) will be used in eFflows4HPC Pillar III (urgent computing for natural hazards). The corresponding workflows proposed by eFlows4HPC would certainly have a beneficial impact on ChEESE services in such a relevant application domain. I wish you every success with the bid, and would be very happy to be kept updated of the development of the project.

Yours sincerely,

Dr. Arnau Folch Duran
PI of the ChEESE Center of Excellence (https://cheese-coe.eu)
e-mail: arnau.folch@bsc.es

To the attention of
Rosa M Badia
Coordinator of the
eFflows4HPC proposal
(objective EuroHPC-02-2019):

Dr.-Ing. Bastian Koller
Nobelstr. 19 • 70569 Stuttgart
Germany
Tel.: +49 (0)711/685-65891
Fax: +49 (0)711/685-65832
eMail:koller@hlrs.de

19.12.2019

Letter of support from the EXCELLERAT Centre of Excellence for the eFflows4HPC proposal

EXCELLERAT is a Center of Excellence that offers cross-cutting support for various engineering sectors, like manufacturing, automotive, energy, aerospace, chemistry, biology and climate, enabling generic application support is a non-trivial task. Within the engineering sector, we see specialisation on expertise, applications development, targeted training and offered hardware.

eFlows4HPC aims to deliver a workflow software stack and additional set of services that will enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The project will integrate existing workflow interfaces, programming models, machine learning and data analytic libraries to provide a uniform platform that enable the exploitation of future large-scale systems. Furthermore, eFlows4HPC will deliver a methodology to deploy and re-use workflows to widen the usage of HPC to new user communities. The project will demonstrate through three application pillars with high industrial and social relevance: manufacturing, climate and urgent computing for natural hazards, how the realization of forthcoming efficient HPC and data centric applications can be developed with new workflow technologies.

The manufacturing pillar deals will focus on the integration of state-of-the-art adaptive solvers with the eFlows4HPC workflow stack management system to allow the digital prototyping of complex manufactured objects, based on the KRATOS General Purpose, Parallel Finite Element solver, which contains basic tools for Full Order and Reduced Order models. The pillar will thus aim at the construction of Digital Twins to be used both in the design and in the lifelong deployment of manufactured objects, as envisaged in the framework of the Industry 4.0 vision. The resulting workflow will be made available to the EXCELLERAT community, both the instance developed for the project and a generic template that could be tailored for specific uses. This will benefit the eFflows4HPC project in its exploitation activities and the EXCELLERAT community that will be benefit from a new type of workflow of their interest.

Dr.-Ing. Bastian Koller
Coordinator of the EXCELLERAT - Centre of Excellence

To: Rosa M. Badia
Coordinator of the eFlows4HPC proposal
Barcelona Supercomputing Center (BSC)

Hamburg, 02. 01. 2020

Subject: Letter of Support to eFlows4HPC

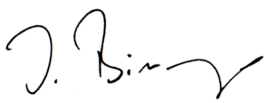
As coordinator of the Center of Excellence for Weather and Climate (ESiWACE2), I express our strong interest and motivation to collaborate with the researchers involved in the proposal entitled *"Enabling dynamic and Intelligent workflows in the future EuroHPC ecosystem – eFlows4HPC"*, submitted to call EuroHPC-02-2019.

The objective of ESIWACE is to prepare the weather and climate community to be able to make use of exascale systems when they become available. This includes extremely compute intensive applications associated with complex and data intensive workflows.

eFlows4HPC intends to deliver a workflow software stack and additional set of services that will enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The project will integrate existing workflow interfaces, programming models, machine learning and data analytic libraries to provide a uniform platform that enable the exploitation of future large-scale systems. The project will demonstrate through three application pillars with high industrial and social relevance: manufacturing, climate and urgent computing for natural hazards, how the realization of forthcoming efficient HPC and data centric applications can be developed with new workflow technologies. The project aims also at delivering methodologies and tools that widens and makes easier the use of workflows by existing and new HPC communities and users.

eFlows4HPC considers the climate workflows one of the key use cases where the research in the project can be applied, devoting the eFlows4HPC Pillar II to this topic. In this pillar, improved workflows for ensemble climate simulations will be developed using some of the state-of-the-art Earth System Models as OpenIFS or FESOM2. In addition, workflows involving post-processing climate data with Machine Learning techniques will also be covered.

We would be happy to be kept updated of the development of the project and consider future collaborations and provide feedback regarding the workflows developed.



Joachim Biercamp
DKRZ, Head of Application Departement
(Coordinator of ESiWACE)

ESiWACE2 has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 823988

THE UNIVERSITY OF CHICAGO
COMPUTER SCIENCE

Ian Foster
Professor

December 15, 2019

Letter of Support

To the attention of **Rosa M Badia**,
Coordinator of the **eFflows4HPC** proposal (objective EuroHPC-02-2019):

I am Professor of Computer Science at the University of Chicago and Director of the Data Science and Learning Division at Argonne National Laboratory. My group develops and applies methods for data-intensive computing in high-performance computing environments, with applications in such domains as life sciences, materials sciences, and scientific instrumentation. We produce workflow tools such as Swift and Parsl, automation tools such as Globus Automate, and data services such as Globus that are used by hundreds of thousands of people worldwide.

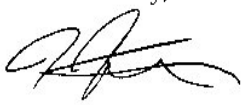
I understand that your proposed eFlows4HPC project intends to deliver a workflow software stack and services that will enable the integration of HPC simulation and modelling with big data analytics and machine learning in scientific and industrial applications. The project will integrate existing workflow interfaces, programming models, machine learning, and data analytics libraries to provide a uniform platform that enables the exploitation of future large-scale systems. It will demonstrate, through three application pillars with high industrial and social relevance: manufacturing, climate, and urgent computing for natural hazards, how the realization of forthcoming efficient HPC and data centric applications can be

5730 S. Ellis Avenue • Chicago, Illinois 60637
Crerar 301 • Telephone: 773-702-3487 • Email: foster@uchicago.edu

developed with new workflow technologies. The project aims also to deliver methodologies and tools that widen and make easier the use of workflows by existing and new HPC communities and users.

Given our strong common interests in big data and science automation, and our past record of fruitful collaboration, I look forward to working with you and other members of the **eFflows4HPC** project once it is funded.

Yours sincerely,



Dr. Ian Foster

Arthur Holly Compton Distinguished Service Professor
Department of Computer Science, The University of Chicago

Senior Scientist and Distinguished Fellow
Director, Data Science and Learning Division, Argonne National Laboratory

Annex B. External Advisory Board Profiles

Dr. Andrey Y. Babeyko is a senior researcher at the GeoForschungsZentrum-Potsdam. Since 2006, responsible for the work-package "Tsunami source modeling", from the German Indonesian Tsunami Early Warning System (GITEWS). Dr Babeyko provides tools and data for modeling of tsunami excitation and propagation, and tools, data and professional expertise for testing and verification of the GITEWS Decision Support System. Has been developing a GUI-based combined earthquake- and tsunami- modeling software for teaching and training of TEWS personnel. Has developed a novel tsunami early warning technique based on real-time GPS arrays. Does research in methodology and tools for probabilistic tsunami hazard assessment, and does research on the modeling effects of phase transformations in oceanic crust and mantle on the stress distribution in the subducting plate.

Elias Cueto is a Professor of continuum and computational mechanics, Universidad de Zaragoza. His research interests include the development of numerical methods for computational mechanics in its broadest sense. He has worked on finite element and meshless methods or model order reduction techniques, with applications on forming process simulation, real-time simulation, haptics, computational surgery, and, more recently, on data-intensive computational mechanics and Augmented/Mixed Reality. Prof. Cueto has been awarded with the J. C. Simó prize of the Spanish Society for Numerical Method in Engineering, SEMNI, the Scientific Prize of the European Scientific Association of Material Forming (ESAFORM) or the O. C. Zienkiewicz of the European Community on Computational Methods in Applied Sciences (ECCOMAS), among other distinctions. He is currently the president of the Spanish Society for Numerical Methods in Engineering (SEMNI).

Prof. Dr. Daniela Jacob is meteorologist and Director of the Climate Service Center Germany (GERICS), a scientific organizational entity of Helmholtz-Zentrum Geesthacht, and visiting professor at Leuphana University, Faculty of Sustainability. She was coordinating lead author of the IPCC Special Report on the impacts of global warming of 1.5°C above preindustrial levels, and one of the leading authors of the IPCC Fifth Assessment Report (Working Group 2). Daniela Jacob is chair of German Committee for Sustainability (DKN), and member of several other committees, as well as Ex-officio member of the 'Earth League', an international alliance of prominent scientists from world-class research institutions. Her main research fields and areas of interest are regional climate modelling and the hydrological cycle. Moreover, Daniela Jacob is Editor-in-Chief of Journal "Climate Services", a new scientific Journal she co-founded with Elsevier. She is member of the European Commission's Mission Board on "Adaptation to Climate Change including Societal Transformation".

Dominique Ragot, Ing, graduated from ENST, Paris, France in 1988. Presently Senior Software Architect and Real-Time Expert at Thales group, has been involved in the design and integration of many advanced software intensive processing systems at Thales including radars, sonar, infrared imaging, satellite and telecommunication systems. For more than 20 years I have been participating in many collaborative projects mainly at European level (Eureka, ITEA, EDA, FP7, H2020). My areas of expertise are at the architecture/application adequation in order to provide the most efficient solution for industrial systems with non-functional constraints such as reliability, safety and security. In this context I have been working in HPC with a special focus on heterogeneity and techniques to manage applications in order to optimize both processing and communications vs. power and QoS constraints. Recent H2020 projects: SAFURE, MANGO.

Annex C. Acronyms table

AI	Artificial Intelligence
ARB	Architecture Review Board
BDEC	Big Data and Extreme-scale Computing
BoF	Birds of a Feather
CA	Consortium Agreement
CaaS	Container as a Service
CC	Cyclomatic Complexity
CF	Climate Forecast
CMIP	Coupled Model Intercomparison Project
CMT	Centroid-Moment-Tensor
CoE	Center of Excellence
CSA	Coordination and Support Action
CUDA	Compute Unified Device Architecture
DA	Data Analytics
DEIM	Discrete Empirical Interpolation Method
DISS	Database of Individual Seismogenic Sources
DL	Deep Learning
EAB	External Advisory Board
ECSW	Energy Conserving Sampling and Weighting
ECV	European Computer Vision library
EDD	European Distributed Deep Learning library
EIM	Empirical Interpolation Method
EPI	European Processor Initiative
ERCC	Emergency Response Coordination Centre
ESGF	Earth System Grid Federation
ESM	Earth System Model
FaaS	Function-as-a-Service
FEM	Finit Element Method
FESOM	Finite-Element/volumE Sea ice-Ocean Model

FOM	Full Order Model
FPGA	Field-Programmable Gate Array
FTRT	Faster Than Real-Time
GMM	Gaussian Micture Models
GPU	Graphics Processing Unit
HDF	Helmholtz Data Federation
HDF5	Hierarchical Data Format v5
HeAT	Helmholtz Analytics Toolkit
HPAC	High Performance Analytics and Computing
HPC	High Performance Computing
HPCWaaS	High Performance Computing Workflow-as-a-Service
HPDA	High Performance Data Analytics
HySEA	Hyperbolic Systems and Efficient Algorithms
IaaS	Infrastructure-as-a-Service
ICT	Information and Communication Technology
IP	Intellectual Property
IPR	Intellectual Property Rights
ISC	International Supercomputing Conference
ITHACA	In real Time Highly Advanced Computational Applications
JSON	JavaScript Object Notation
JU	Joint Undertaking
LoC	Lines of Code
ML	Machine Learning
MLMC	Multilevel Monte Carlo method
MMG	Multidisciplinary Meshing Library
MoV	Means of Verification
MPI	Message Passing Interface
MRL	Market Readiness Level
NVRAM	Non-Volatile Read Access Memory
OpenIFS	Open Integrated Forecasting System

ParSODA	Parallel Social Data Analytic
PFV	Peak Ground Velocity
PGA	Peak Ground Acceleration
PGM	Peak Ground Motions
PI	Principal Investigator
PSO	Pillar-specific Scientific objective
PTF	Probabilistic Tsunami Forecasting
PTHA	Probabilistic Tsunami Hazard Assessment
PyCOMPSs	Python Comp Superscalar
QoS	Quality of Service
RDA	Research Data Alliance
REST	Representational State Transfer
ROM	Reduced Order Model
RSN	Rete Sismica Nazionale
SC	Supercomputing Conference
SIO	Societal and industrial objective
SSD	Solid State Memory
STO	Scientific and Technological objective
SVD	Sing**le** Value Decomposition
SWOT	Strengths-Weaknesses-Opportunities-Threats Analysis
TC	Tropical Cyclone
TEWS	Tsunami Early Warning System
TOSCA	Topology and Orchestration Specification for Cloud Applications
TPU	Tensor Processing Unit
TRL	Technology Readiness Level
USGS	United States Geological Survey
VOL	HDF5 Virtual Object Layer
Yorc	Ystia Orchestrator

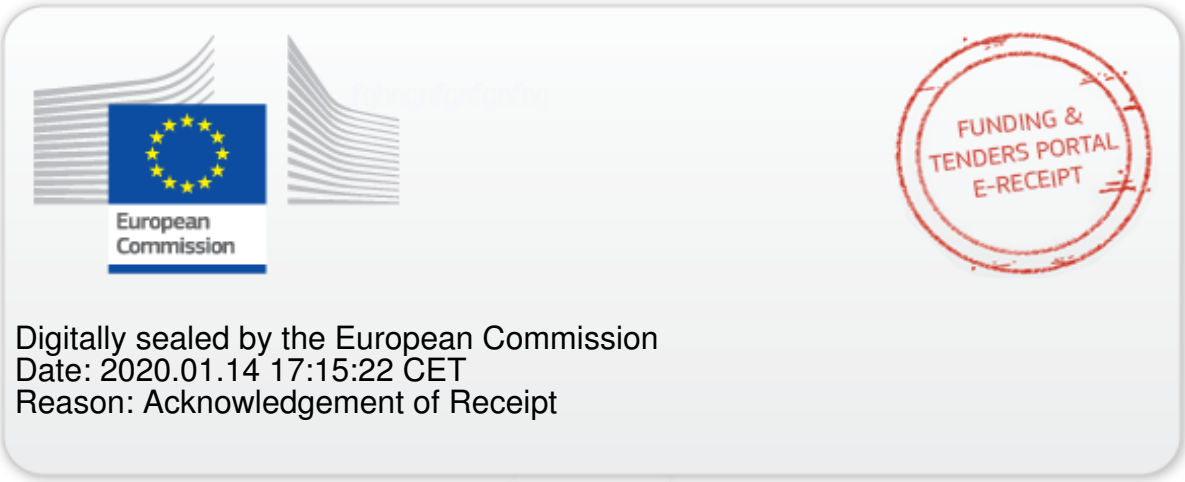
Annex D. Table summarizing roles and partners expertise

Partner	Main Role	Expertise Relevance for the Success of the Project
BSC - WDC	Coordinator, technical leader, WP1 leader, project management & dissemination manager. Technology provider. Supercomputing center	Experts in workflow environments, programming models for heterogeneous systems, persistent storage and machine learning. Contributes with the PyCOMPSs workflow environment, with the dataClay and Hecuba/Qbeast persistent storage libraries and with the dislib machine learning library.
BSC - CASE	Pillar III leader (WP6). Workflow provider.	Experts in natural hazards and design/development of HPC modeling workflows. Contributes with the UCIS4EQ workflow.
BSC - Earth	Contributor to WP5. Workflow provider. Data provider. End user.	Experts in HPC applied to Earth System Models and designing tools and workflows to run coupled climate models. Support from Climate Prediction group for climate science and also end-users of the developments of the project. BSC Earth is also responsible for a Tier-2 ESGF data node.
CIMNE	Pillar I leader (WP4). Workflow provider.	CIMNE has a long-standing expertise in high performance solver development. It leads the development of the general purpose code Kratos,. Within the project it will be in charge of integrating enhanced solution technologies in the Flows4HPC framework.
FZJ	WP2 leader. Technology provider. Supercomputing Center	FZJ has a long tradition as a world-leading HPC center. Experts in architecting and implementing (distributed) data intensive applications, Cloud computing, and emerging architectures. Experts in scientific software including machine learning.
UPV	WP3 leader. Technology provider.	Expert on acceleration of algorithms on FPGAs and other accelerators. Machine learning expert.
ATOS	WP7 leader. Exploitation leader. Technology provider.	Expert in HPC, AI and Cloud technologies. Contributes to high-level workflow management leveraging the YSTIA/YORC TOSCA based orchestrator. Contributes to AI models management, leveraging its expertise and tools suite for AI/HPC. Experts on innovation and technology transfer.
Dtok Lab	Technology provider	Expert in machine learning, data mining algorithms and services, scientific workflow management, and Cloud computing.
CMCC	Pillar II leader (WP5). Technology provider. Workflow provider. Data provider. End user.	Scientific expertise on Earth System Models and climate modelling workflows, climate simulations and predictions, extreme events (e.g. tropical cyclones). CMCC runs two ESGF data nodes for dissemination of CMIP5 and CMIP6 data. Expertise on HPDA, big data, ML applied to climate use case and data science software/applications. Contributes with Ophidia to HPDA workflows/applications and with the CMCC-CM2 coupled model and ESM workflows in Pillar II.
INRIA	Technology provider Workflow provider	Expert in high performance adaptive and embedded solvers, error estimation and mesh adaptation tools
SISSA	Technology provider Workflow provider	Expert in Reduced Order Modelling (ROM), parametric formulation, inverse problems, numerical analysis, scientific computing, uncertainty quantification.

PSNC	Supercomputing center	Expertise in policies for Urgent Computing.
UMA	Technology provider Workflow provider	Expertise in numerical tsunami modelling as code developers. Large experience in the implementation of reliable and efficient massively parallel numerical codes for geophysical flows using self-designed methods and algorithms. A deep knowledge of the codes and algorithms implemented is crucial for project success.
INGV	Workflow provider. Data provider. End user	Experts in Tsunami and earthquake
AWI	Technology provider Workflow provider End user	AWI has an expertise in ocean and climate model development with emphasis on high resolution simulations and participate in CMIP6 activity. It develop in-house the next generation global ocean model that use unstructured meshes (FESOM2). Within the project AWI will provide coupled climate model OpenIFS/FESOM2 for Pillar 2 use case and participate in it's co-development. AWI will lead collection of requirements for Pillar II workflow and it's final evaluation.
ETHZ	Workflow provider	HPC numerical wave propagation codes provider, Earthquakes' workflows expert
Siemens	Technology provider. End user	Expert in model order reduction as a key technology to enable Digital Twins
NGI	Workflow provider	Expertise in numerical tsunami modelling covering a wide range of applications, from probabilistic analysis to complex source modelling. The experience from ChEESE on developing methods and workflows for tsunami HPC and urgent computing is particularly relevant for Pillar III.

Annex E. SWOT analysis

Strengths - Partners bring software tools, libraries and expertise in the different topics of workflows for HPC, big data and artificial intelligence with technological baseline with high TRL - Range of software solutions covering similar needs - And expertise and tools on the different pillars' application areas. - Setup for a European Workflow platform that enables the convergence of HPC, HPDA and AI is optimal - Partners with previous collaborations between them - Strong proposal of the eFlows4HPC software stack for the convergence of HPC, HPDA and AI with energy efficiency - HPCWaaS proposed methodology to ease development, usage and sharing of workflows - Partners with access to top HPC infrastructures - High potential industrial and societal impact - Proposal that develops and maintains a world-class HPC ecosystem in the Union - General-purpose solution not limited to a specific area of knowledge	**Weaknesses** - Possible staffing issues due to high demand on IT personnel - Integration overheads: Integrating components from different institutions could require important efforts and developments.
Opportunities - Strategic situation of High-Performance Computing and Artificial Intelligence (currently included in the portfolio of the Commissioner for Internal Market Ms Sylvie Goulard). - The convergence of HPC, HPDA and AI considered as strategic by multiple international initiatives like the BDEC. - Creation of the EuroHPC JU and its roadmap for Exascale - The availability of European workflow platforms that enable the integration of HPC, HPDA and AI, capable of benefiting from Exascale systems is not only optimal, but strategic. - Opportunity to achieve an integrated workflow for the simulation-based Digital Twins, which are the basis of "Industry 4.0". - Related projects with good links to the partners - Participation of consortium members in related CoEs	**Threats** - Lack of adoption from user communities - Competitors: Other current workflow engines



Digitally sealed by the European Commission
Date: 2020.01.14 17:15:22 CET
Reason: Acknowledgement of Receipt

This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Funding & Tenders Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Funding & Tenders Portal.

(https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq)