D7.3 Training Concept

Document Identification			
Status	Final	**Due Date**	31/03/2020
Version	1.0	**Submission Date**	08/04/2020

Related WP	WP7	**Document Reference**	D7.3
Related Deliverable(s)	D4.2, D5.1, D5.3, D7.2	**Dissemination Level (*)**	PU
Lead Participant	USTUTT	**Lead Author**	Lorenzo Zanon
Contributors	BUL, DIA, ECMWF, KNOW, PLUS, PSNC, SZE	**Reviewers**	Alexandra Gens (DIA) John Hanley (ECMWF)

Keywords:
Training, Education, Global Challenges (GC), High-Performance Computing (HPC), High-Performance Data Analytics (HPDA), Best Practice, Training Concepts

Document Information

List of Contributors	
Name	Partner
Derek Groen	BUL
Alexandra Gens	DIA
John Hanley	ECMWF
Milana Vuckovic	ECMWF
Konstantinos Nikas	ICCS
Dimitrios Tsoumakos	ICCS
Mark Kröll	KNOW
Robert Elsässer	PLUS
Marcin Lawenda	PSNC
Zoltán Horváth	SZE
Sergiy Gogolenko	USTUTT
Dennis Hoppe	USTUTT
Natalie Lewandowski	USTUTT
Anna Mack	USTUTT
Dineshkumar Rajagopal	USTUTT
Lorenzo Zanon	USTUTT

Document History			
Version	Date	Change editors	Changes
0.1	11/02/2020	L. Zanon (USTUTT)	Created initial document, Table of contents.
0.2	05/03/2020	ALL	Draft.
0.3	13/03/2020	ALL	Complete draft for QA.
0.4	17/03/2020	L. Zanon (USTUTT)	Revisited draft at QA (first round of internal review), following comments by A. Gens (DIA) and D. Hoppe (USTUTT).

Document History			
Version	Date	Change editors	Changes
0.5	25/03/2020	L. Zanon (USTUTT)	Revisited draft at QA (first round of internal review), following comments by J. Hanley (ECMWF).
1.0	29/03/2020	L. Zanon (USTUTT)	Final version (no changes needed after second review).

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Lorenzo Zanon (USTUTT)	29/03/2020
Quality manager	Marcin Lawenda (PSNC)	30/03/2020
Project Coordinator	Francisco Javier Nieto (ATOS)	08/04/2020

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
ABMS	Agent-based modeling and simulation
AI	Artificial Intelligence
API	Application Programming Interface
ASD	Agile Software Development
ATOS	ATOS SPAIN SA
BUL	BRUNEL UNIVERSITY LONDON
CoE	Centre of Excellence
CSA	Coordination and Support Action
DIA	DIALOGIK GEMEINNUETZIGE GESELLSCHAFT FUER KOMMUNIKATIONS- UND KOOPERATIONSFORSCHUNG mbH
DL / ML	Deep Learning / Machine Learning
DoA	Description of the Action
EC	European Commission

Abbreviation / acronym	Description
ECM	Enterprise content management
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
GC	Global Challenges
GPU	Graphics processing unit
HLRS	High-Performance Computing Center Stuttgart
HPC	High Performance Computing
HPDA	High Performance Data Analysis
ICCS	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS
KNOW	KNOW-CENTER GMBH RESEARCH CENTER FOR DATA-DRIVEN BUSINESS & BIG DATA ANALYTICS
KPI	Key Performance Indicator
PLUS	PARIS-LODRON-UNIVERSITAT SALZBURG
PSNC	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK
SNA	Social Network Analysis
SZE	SZECHENYI ISTVAN UNIVERSITY
UAP	Urban Air Pollution
USTUTT	UNIVERSITAET STUTTGART
WP	Work Package
Y1, Y2,…	Year one, two, … from the beginning of the project (December 2018)

Executive Summary

This deliverable describes the efforts conducted by HiDALGO Task 7.3 (T7.3) in bringing together the Global Challenges (GC) and High Performance Computing (HPC) communities by providing a training curriculum to GC scientists. This document shows how specific HPC requirements arising when dealing with GC can be collected in order to come up with such a curriculum. Additionally to the collection of requirements, the quality of the HiDALGO curriculum depends on both didactic and technical best practices in training. At the same time, training events are opportunities to exchange among the different communities, and therefore to collect feedback on the technical pilots of the project. In short, this document describes current and planned events, new requirements, best practices, and collected feedback.

HPC requirements in GC arise naturally from the development of the use cases. It is shown in this document how these requirements are mostly being tackled through the interaction among the technical teams in WPs 3 to 6. Further requirements arising from the CoE stakeholders will be identified in conjunction with the tasks focused on outreach (e.g., T7.2, T4.5, and WP2). These requirements represent a motivation for HPC and GC communities to come together at "innovation workshops", which have been organised and planned in tight cooperation with T7.2. This document analyses their outcome – in terms of educational results, collected feedback, and acquired best practices. An even stronger educational input is given by the introduction of the CoE's technical expertise in the established training curriculum of the partners. HiDALGO contributions within face-to-face courses are considered, as well as training online tools such as webinars and Massive Open Online Courses (MOOCs). Key Performance Indicators (KPIs) to measure and boost the outcome of this task in year 2 (Y2) and Y3 are proposed at the end of this deliverable.

Drawing up a training concept highlighted the necessity of several interactions between T7.3 and other subjects:

- with T7.2 for the organisation and planning of future events;
- with the technical tasks in order not to miss training opportunities in conjunction with technical developments;
- with other CoEs to share the management of (training) events and increase the learning outcomes.

1 Introduction

1.1 Purpose of the document

This deliverable analyses the current state-of-the-art of HiDALGO training activities and innovation workshops and defines the necessary roadmap for future events. The aim of such events is to raise mutual awareness amongst the GC and HPC/HPDA communities. On the other hand, strategies are laid out to define a compendium to address the requirements in HPC and HPDA expertise for GC scientists. Other equally important aspects of T7.3 that are described in this document include the collection of feedback on the pilots from stakeholders and the academic community, interaction with other CoEs, and the identification of best practices in education and training.

1.2 Relation to other project work

The activities and concepts described in this document are strictly intertwined with the purposes of T7.2 (Event Management and Collaboration). In this respect, D7.3 expands and updates D7.2 [33] from the point of view of training. Organising (training) events and performing requirement analyses is also related to the work carried out by T7.1 (Awareness Creation and Community Support) and T7.2, but also T4.5 (Future applications) and WP2 (Business models and sustainability). Additionally, the set-up of technical tools for training best practice depends on the activity of WP5, dedicated to the establishment of the HiDALGO Portal (see D5.3 [31]). In particular, T5.2 (HiDALGO Support Infrastructure) provides customer support for clarifying queries related to training, while T5.3 (HiDALGO Portal) implements a training service in the HiDALGO Web Portal for disseminating training material.

1.3 Structure of the document

This document is structured into seven major Sections, including Document Information, Introduction (Section 1), and Conclusion (Section 7):

- Section 2 describes strategies to promote the interaction between the HPC/HPDA and GC communities, which will ultimately lead to a training curriculum.
- Section 3 details the joint effort of T7.2 and T7.3 to enable HiDALGO to play a proactive part at “innovation workshops”.
- Section 4 focuses on purely educational activities, especially those that can be integrated in the established training curriculum of the project partners.

- Section 5 sums up suggestions on how to improve upcoming training activities, also by enhancing the technical aspects and by fostering the participation of attendees with limited means.
- Section 6 concludes the core of the document by presenting the KPIs for T7.3, along with a few collateral initiatives.

2 A Curriculum for Global Challenge Scientists

One of the main goals of T7.3 is to create a training curriculum for HPC technology leaders. Such scientists and analysts should be given the tools to tackle GC from an intertwined HPC and HPDA perspective.

For instance, the partners' expertise in terms of HPDA (Big Data, Machine/Deep Learning, Artificial Intelligence, etc.) will become part of the training curriculum, also exploiting HiDALGO-specific contributions (Section 4.3). In order to provide a consistent curriculum though, this task strives to offer training that is as specific and tailored as possible. To do so, the gaps in HPC and HPDA knowledge of experts working on GC must be identified, mainly in the two ways described in the following sections.

2.1 HPC and HPDA Requirements from the Pilots

GC simulations have given rise to a number of new requirements of HPC environments, which have been elaborated within the HiDALGO use cases. While the reader can consult [22] for a detailed description of these requirements, a summary follows:

- *Support for Python and R*: HiDALGO pilots extensively use the Python and R ecosystems for data analytics and visualization in HPC. In addition, Python is often used for fast prototyping of the simulated models (e.g. in the Migration pilot), as well as for certain data management operations in the UAP and Social Network Analysis (SNA) pilots.
- *Support for agent-based modeling paradigm*: In contrast to traditional HPC applications from engineering and science domains, GC applications naturally contain parts related to social simulations. Agent-based modeling and simulation (ABMS) is a fundamental pillar of computational social sciences and the most popular instrument for micro-level social simulations. Thus, HPC environments for GC must include a wide range of HPC-compliant ABMS instruments, such as general-purpose ABMS frameworks (RepastHPC, AMOS, etc.) along with problem-oriented ABMS tools (SUMO, Flee, etc.).
- *Coupling with both external static and streaming data sources*: HPC environments typically operate on static data, which is nowadays already feasible on efficient parallel distributed file systems such as Lustre, BeeGFS, or IBM Spectrum Scale. However, being intended to reflect current and upcoming social situations for policymaking, GC simulations extensively use real-time data. These originate from external sources including streaming data from physical sensors, social networks, mobile operators, etc. In particular, streaming traffic sensor data are coupled with UAP data, while external telecommunication data with those of the Migration and SNA pilots. Finally, the integration of the streaming data into HPC environments can also contribute to expanding the use of HPC systems to the domains of Internet of Things and Industry 4.0.

- *Coupling simulations across data-centres*: This requirement stems from the necessity to couple simulations and exchange data across data centres in GC simulation scenarios. Such a demand may arise for various reasons, e.g. re-using expensive simulation results in several separate use cases, the inability to share massive amounts of data, and issues related to licensing the data or simulation software at different data centres. In the HiDALGO project, we couple weather forecast simulations and meteorological data (ECMWF) with CFD and ABMS simulations (Migration and UAP pilots) on the resources of both USTUTT and PSNC.

One of the main goals of HiDALGO is to bring together GC simulations and Big Data Analytics processing in order to enhance and accelerate performance, quality, and response times. Given the current state of the HiDALGO pilots, an overview of the emerged HPDA requirements follows (see also D6.4 [32]):

- *Support for high-performance, general-purpose analytics runtime(s)*: Pilots have registered and to-date required method implementations over huge numerical data sources (UAP: standard matrix operations, aggregations, etc.), graph-analytics operations (Migration Pilot: link detection, clustering, etc.), and streaming data crunching and coupling (for all the three pilots, using sensor traffic and weather data).
- *Support/ability of coupling the aforementioned runtime(s) with high-performance datastores or distributed file systems* such as HDFS, Lustre, CKAN, relational/main memory databases, Apache Avro/Parquet, etc. (for all the three pilots).

While choosing a single engine for all tasks and data has proven to be highly inefficient, a combination of tools and programming languages has been preferred, namely:

- *Support for Apache Spark™ [23]*, a general purpose main memory computing runtime that supports both batch and stream processing, with extensive graph and Machine Learning (ML) operations available. Spark™ can easily connect to a variety of data sources and tools and be utilized under Java and Python (which is important to many data scientists). Similarly, Spark™ can be used with notebooks and R.
- *Support for Apache Kafka®* [24] (for more specialized stream processing), a distributed streaming platform that stores streams of records in a fault-tolerant durable way and processes them as they occur.

2.2 Stakeholders' needs

Stakeholders' needs in terms of both HPC and HPDA knowledge will be detailed through a requirement analysis, which will be initiated in Q1/Q2 2020. Such an analysis will be achieved jointly with T7.1 (Awareness Creation and Community Support) and T7.2 (Event Management and Collaboration). Once the stakeholder profiles have been analysed and HiDALGO offerings have been shaped in more detail by WP2, potential stakeholders will be contacted to initiate a collaboration with our project. On this occasion, training needs will also be investigated with the help of partners technically involved in the use cases, such as BUL, KNOW, PLUS, PSNC,

SZE, and USTUTT. Other possible approaches include adding a survey on the CoE webpage [11], or circulating printed copies at the innovation workshops (Section 3).

In such a multi-purpose survey, an inquiry about training will have the form:

HiDALGO can offer training in various domains: Which specific topics would be most interesting for you?

- ☐ *Introduction to HPC: A beginners' guide.*
- ☐ *Introduction to Global System Sciences and its Challenges.*
- ☐ *High Performance Computing vs. Cloud. When to use what?*
- ☐ *Developing for HPC using parallelization techniques.*
- ☐ *Code Optimization for HPC.*
- ☐ *Applying Artificial Intelligence to get new insights from my data.*
- ☐ *Introduction to specific application areas:*
    - *Migration*
    - *Urban Air Pollution*
    - *Spread of Messages in Social Networks*
- ☐ *Other: [Please specify]*

Based on the existing and upcoming requirements, T7.3 developed a curriculum consisting, on the one hand, of innovation workshops, detailed in Section 3.1 (past) and Section 3.3 (upcoming). While innovation workshops have further goals beyond training (collection of feedback, dissemination and outreach, etc.), the activities described in Section 4.1 (past) and Section 4.3 (upcoming) are focused on education and are a fundamental part of the curriculum. An overview of all performed and planned activities building up the HiDALGO curriculum can be found in Table 2 in the Annexes.

One of the valuable sources of training are the tools developed within the case studies by the HiDALGO scientific experts. As an example, the simulation of the Urban Air Pollution (UAP) pilot with OpenFOAM® was exhibited at a hands-on session at the HiPEAC conference (Section 3.1). OpenFOAM® [26] is an open-source code and powerful framework for solving problems in the field of CFD. Training sources are one important aspect which defines the curriculum quality, which is assessed and continuously verified through the analysis and improvement of didactic and technical tools. These best practices are summarized in Section 5 and together with the proposed curriculum make up the HiDALGO training concept.

3 Innovation Workshops

HiDALGO organises and participates in conferences, workshops, and other scientific events, referred to in the Description of the Action (DoA) with the umbrella-term “innovation workshops”. They have the multifold purpose of collecting feedback on HiDALGO’s activities (Section 3.2), identifying best practices in education and training (Section 5), and performing education activities to bring the centre’s expertise to the users.

This endeavour is strictly intertwined with the objectives of T7.2 (see D7.2 [33]): Workshops and sessions at conferences are planned with the shared goal of driving mutual GC and HPC/HPDA awareness and educating the participants. HiDALGO strives to organise some of these events jointly with other projects, e.g. with EoCoE-II [14] (first contact currently planned), ESiWACE-2 [15] (established contact regarding the use cases), or EXCELLERAT [13] (see Sections 3.1, 3.2, 3.3, 4.1).

3.1 Activities performed

HiDALGO participated at the HiPEAC conference [1] (European Network on High Performance and Embedded Architecture and Compilation, Bologna, Italy, 20-22 January 2020) with a half-day workshop [2] (HPC and Big Data Technologies for Global Systems Interactive Workshop and Hands-on Session, see Figure 1 and Figure 2). This included three talks together with a demonstration and hands-on tutorial of the UAP pilot (overview of the challenge, simulation with OpenFOAM® in HPC, basics of blueprints and web interfaces, deploying and executing jobs in HPC). The session was concluded with a discussion between the project members and the audience (about 35 participants). This event was targeted to HPC experts from academia, PhD students, and industry representatives.

Figure 1. HiPEAC Workshop – Social Network pilot presentation

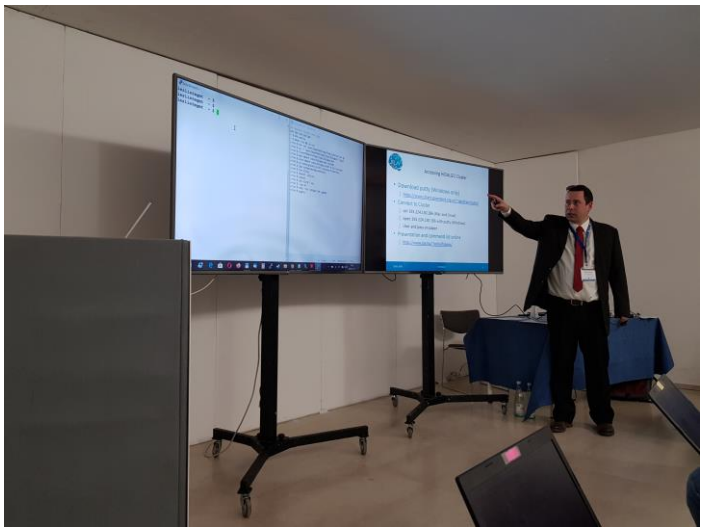

Figure 2. HiPEAC Workshop – UAP pilot presentation

During the workshop, there has been a constructive interaction with the CoE EXCELLERAT [13], which is seeking new engineering applications to boost towards higher scalability and performance (see e.g. EXCELLERAT D5.1 [34] and D5.3 [35]). Within the UAP case study, HiDALGO could benefit from the EXCELLERAT flagship code NEK5000 [25], designed to solve incompressible Navier-Stokes flow simulations at low Mach number. Preliminary tests on the urban air flow have been conducted at USTUTT. Based on these, it can be inferred that NEK5000 achieves better results than OpenFOAM® (currently in use), in terms of speed and scalability.

The audience also showed interest in the SNA pilot, in particular in the Artificial Intelligence (AI) technologies used in there, leading to a discussion with members of the project about possible ways on how to cluster social networks.

3.2 Feedback on the CoE's activity

One of the main goals of innovation workshops is collecting comprehensive feedback on the centre's activities. Such a feedback regards the activities within WP7, in particular training best practices (see Section 5) and cooperation with other CoEs (mentioned at each event description in this document). On the other hand, the project can benefit from feedback from partners and stakeholders on the technical pilots. This section describes the feedback obtained at the innovation workshop HiPEAC, but also at three educational tasks (the Flee Workshop, and the OpenProject and Git webinars, see Section 4.1).

3.2.1 HiPEAC event

At the HiPEAC event [1] (cf. Section 3.1), all pilot applications have been presented. In particular, a hands-on session concerning the UAP pilot with OpenFOAM® [26] lead to a possible collaboration between HiDALGO and EXCELLERAT. Another fruitful potential collaboration could rise between scientists from the University of Pisa and HiDALGO w.r.t. the SNA pilot.

Positive feedback regarding the UAP has also been provided by industrial stakeholders who approached SZE personnel after a UAP presentation with the intention of further contact.

3.2.2 First Flee Workshop

Preliminary education tasks, such as the First Flee Workshop (see Section 4.1) already allowed for the collection of feedback on the Migration case study. For instance, participants in the migration simulation workshop in Adama (Ethiopia) stressed the importance of modelling "internally displaced people" as well as refugees, and also noted the wide range of of modes of transport commonly adopted by displaced people in South Sudan. In addition, positive feedback has been received from several humanitarian organizations, including the UNHCR and the International Organization of Migration [38], which are currently being involved in the HiDALGO efforts to bolster the Migration case study.

3.2.3 OpenProject and Git webinars

Technical task assignments and project progress are tracked easily by OpenProject to improve the clarity between multiple partner contributions and enhance the cooperation among the partners. After the OpenProject webinar (Section 4.1) all the WP leaders created sub-projects within the HiDALGO main project to manage their activities by following Agile Software Development (ASD) principles (Figure 3 and Figure 4).

After the Git webinar (Section 4.1), a central repository has been created for all the HiDALGO tools and technologies to manage the corresponding code (Figure 5). HiDALGO developers are now actively using the Git repository to manage their source code.

A questionnaire has been additionally circulated to investigate users' feedback about these two tools. Addressed issues include problems experienced with the tools before the webinar, those addressed within the webinar, whether these tools have been helpful, or their usage has improved after the webinar. Currently, replies are still being collected.

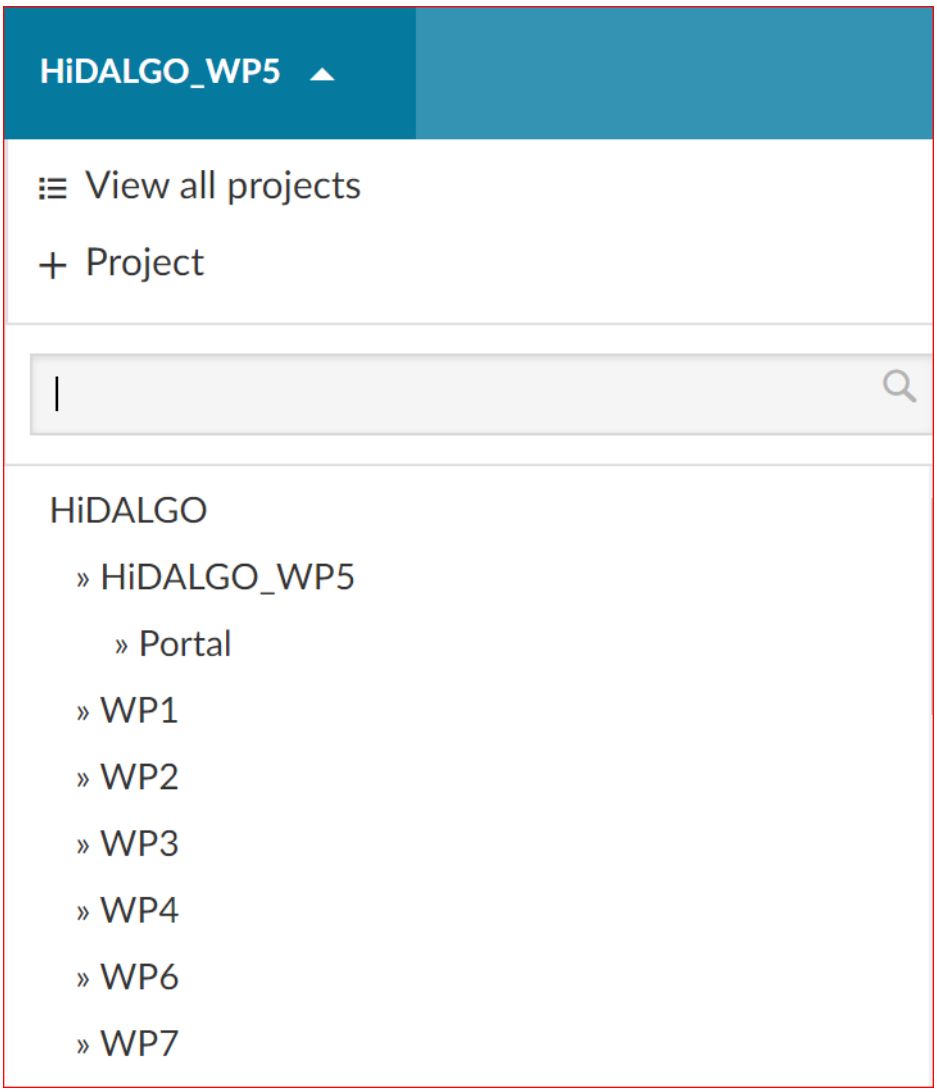

Figure 3. OpenProject: All the WP sub-projects are created and managed by the WP leaders.

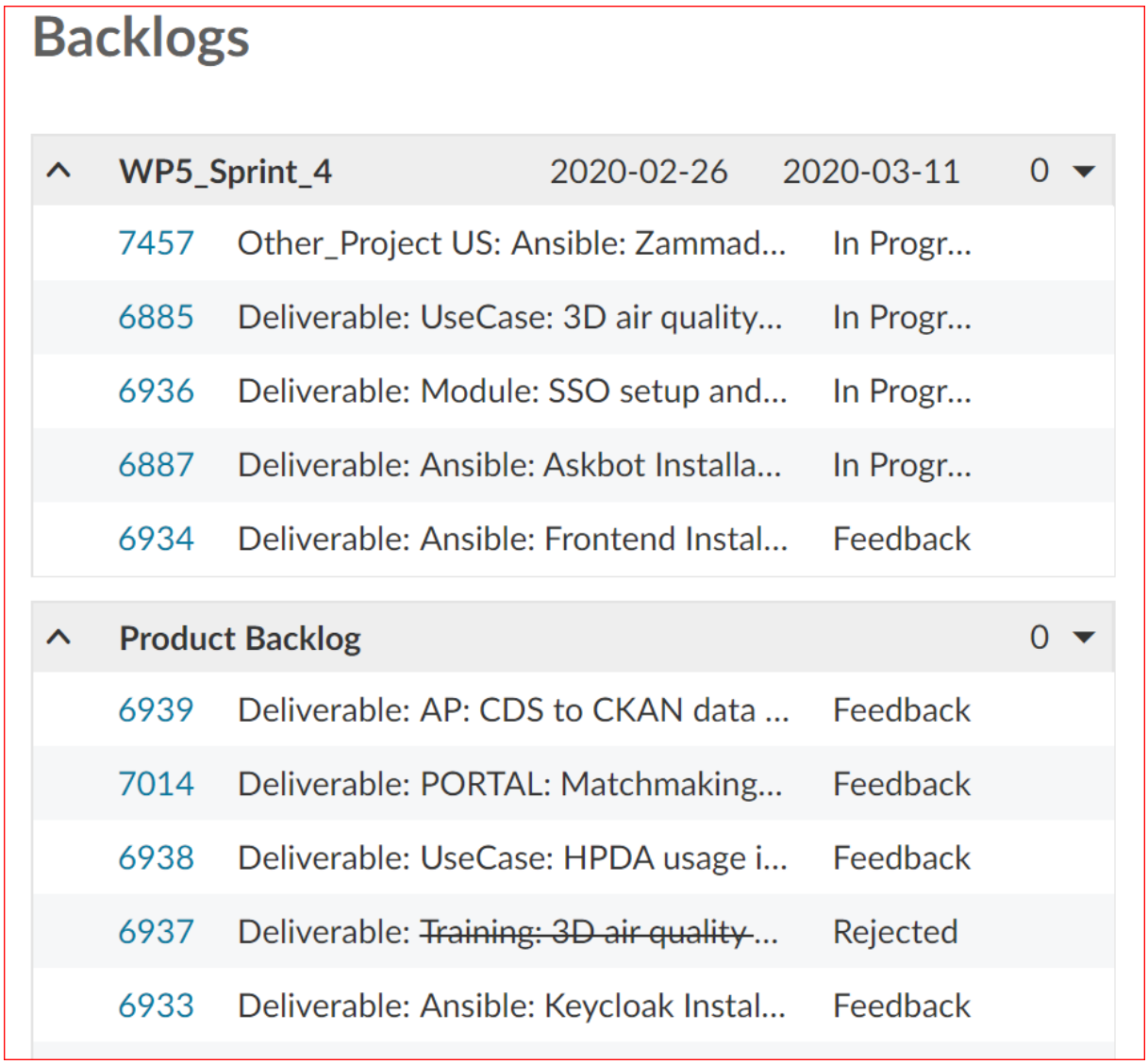

Figure 4. OpenProject: WP5 backlog with the current task information.

The Figure 4 depicts WP5 backlog in the OpenProject with the current task information on WP5_Sprint_4 (sprint of WP5 from 26-02-2020 to 11-03-2020), and the Product Backlog (upcoming tasks for the future sprints). Sprint is an ASD term, corresponding to the planned amount of work that has to be completed and made ready for review by the team within a given amount of time.

Figure 5. A Git project repository has been created for HiDALGO Portal backend activities and partners' contributions.

3.3 Planned activities (roadmap)

In year two (Y2), HiDALGO has been invited to the symposium "Global Challenges: Risk and Complexity" (USTUTT, September 2020). A lecture on the UAP has been already given in the 2019 edition of this event [4]. Thus, it is foreseen that HiDALGO's contribution in 2020 will comprise of a presentation and a workshop showing approaches and intermediate results also from other case studies.

Several conferences have been additionally identified where HiDALGO could propose a workshop or give presentations (see D7.2 [33]). A careful selection is being carried out jointly by T7.2 and T7.3, taking into account impact and topic coherence, but also organisational overhead and logistics. Conferences have the great advantage of a significant impact on the broad scientific community at relatively reasonable cost. Nevertheless, also different kinds of events will have to be considered by T7.2, to additionally boost HiDALGO's outreach to industry. Fairs represent a good start (e.g., the Forum TERATEC 2020 [5]). Feedback about the needs and requests arising from industry is already being collected by WP2 through the FocusCoE [12] network via an online survey.

HiDALGO will have a dedicated session at the workshop on Multiscale Modelling and Simulation [7] within the International Conference on Computational Science [6] (Amsterdam, The Netherlands, 3-5 June 2020). The main topic will be the technical improvement of the CoE's pilots, with a strong focus on coupled simulations. The CoE submitted five talks, as well as a full paper, all of which have been accepted.

HiDALGO will also be present with an overview poster at ISC High Performance 2020 [8] (Frankfurt am Main, Germany, 21-25 June 2020). It would be a great opportunity for the project to give a hands-on tutorial at the 2021 edition of this conference.

Other conferences where HiDALGO might contribute with presentations or posters include the International Conference on Principles and Practice of Multi-Agent Systems (PRIMA) [10], and the Social Simulation Conference [9], where a broad range of topics are covered (among them, ABMS, social networks, traffic simulation, and social impact of climate changes).

Different types of workshops and presentations are additionally considered, e.g. in conjunction with PRACE events [16], in order to maximize the attendance. A possibility would be “teaser workshops” with one-hour presentations for each case study, and hands-on highlighting only a specific technical aspect. This format would have the advantage of presenting the project as a whole (similarly to the poster at the ISC 2020).

Multi-day workshops are taken into account, possibly in cooperation with other CoEs or HPC centres, or as back-to-back initiatives to other established events. In a different fashion compared to conferences, these would be occasions to raise awareness for GC and to build a GC forum. These events should encourage people interested in GC to network and exchange different perspectives towards creating a “global system sciences” community. The CoE would make use of such longer events also to increase the training effort, teaching the use of HPC, HPDA, and AI for GC.

In this regard, WP7 is planning a workshop organised and initiated by HiDALGO to bring together experts working on different GC, which, if approved, will become an Y3 highlight. Additionally, a joint “winter school” in Y2 or early Y3 is in preparation with EXCELLERAT [13] and foresees the involvement of other CoEs (ChEESE [17], FocusCoE [12]). The (preliminary) focus lies on specific HPC topics (load balancing, in situ and remote visualization, GPU code porting). HiDALGO could contribute with technical input from the use cases (WP3 to 6) and with an overview of the internal services originating from the Portal (WP5).

4 Educational Tasks

The goal of educational tasks is to provide training on HPC methods and HPDA, and on how to use AI for tackling GC. Such a knowledge transfer should happen in connection with the growing expertise collected through the case studies. The focus in Y1 and Y2 has been on the Migration and UAP case studies, while the SNA pilot still needs further technical development before such an exercise can be performed. Towards the end of the project, it is envisaged to offer complementary training resulting from the optimized codes.

Educational events are also an opportunity to cooperate with other CoEs to offer (technical) training on methods, as it has been framed by T7.2 (see D7.2 [33]).

4.1 Activities performed

Even though the scheduled start for T7.3 was M12, the Migration case study has been deemed mature enough for a first training event already in M8. HiDALGO organized the First Flee Workshop [3] (University of Adama, Ethiopia, 15-19 July 2019) with the additional involvement from GIZ [37]. This workshop focused on education, covering GC-specific topics (ABMS for addressing GC, modeling of (forced) migration), general core simulation development skills, as well as an introduction to parallel programming and HPC.

For the wider interest of WP7, this workshop allowed for the successful dissemination of the Migration case study by presenting it directly to academic and humanitarian stakeholders in Ethiopia as well as to agricultural agencies. Chances are strong of further collaboration with the University of Adama.

Even though in-class events possess an incomparable effectiveness in know-how transfer, webinars and online activities are capable of reaching a wider audience with considerably simplified participation, also in case of reduced mobility. They are therefore an efficient training tool as well. Moreover, basing e.g. on the assessment performed within EXCELLERAT[1], the method of webinars and online courses is under-represented. At the same time, the "mix-and-match" is recognized among the CoEs as a successful training approach (see e.g. FocusCoE D4.3 [36]). That is, in-class and online events should complement each other, the latter ideally involving interaction among participants and training providers. The current and planned HiDALGO offering in terms of webinars, including presentations, hands-on sessions and discussions, is a promising start in this direction (see also Section 4.3).

[1] Assessment of trainings and workshops of EXCELLERAT [13] Partners, February 2019, to be repeated in Q1 2020. See EXCELLERAT D5.1 [34] and D5.3 [35].

Regarding project-internal webinars, HiDALGO solutions and tools used within the project have been presented at an early stage (before the start of T7.3). Two HiDALGO webinars have been hosted by USTUTT in cooperation with EXCELLERAT [13]:

- Introductory Webinar on OpenProject [18] (2 August 2019) organised by Patrick Vogler and Dineshkumar Rajagopal, with 13 attendees (three female, one from industry, 12 from academia). Goal of OpenProject is to manage all the technical activities within the WPs following the principles of ASD.
- Best Practice Guide for Git & Jenkins [19] (31 October 2019) organised by Patrick Vogler and Dineshkumar Rajagopal, with six attendees (one from industry, five from academia). Git, hosted at USTUTT, is a tool to safely manage the source code of the project.

More information about attendance and feedback analysis on past activities can be found in Section 6.

4.2 Integration of HiDALGO-specific training into existent training courses

The integration of HiDALGO-specific training sessions within the partners' established curricula of workshops and courses is particularly efficient. In fact, among the advantages of this approach, we can mention the wide audience, the reduced organisational overhead, and a fruitful contribution to existing events with HiDALGO original content.

ECMWF, ICCS, KNOW, PSNC, SZE, and USTUTT plan to provide HiDALGO-specific training starting in Y2 or Y3, the reader is therefore referred to Section 4.3. A noticeable effort has been already conducted by PSNC in terms of infrastructure modernization for training purposes.

At PSNC, a special infrastructure has been designed to present the capabilities of the tools implemented within the project activities. This will also allow to facilitate the necessary training and demonstrations for users and potential customers. On such dedicated servers, resources will be available during training activities with ease of access for the participants. In this fashion, limitations due to security and accessibility of HPC systems can be circumvented, and long waiting times due to queueing systems can be avoided. A set of Virtual Machines have been set up at PSNC composed of a sufficient number of CPU cores (128) and memory (128GB) to effectively carry out presentation activities. All applications, corresponding libraries, and tools necessary to demonstrate the pilots' functionalities were deployed and configured on these machines.

4.3 Planned activities (roadmap)

This section provides a roadmap of training activities by the HiDALGO partners ECMWF, ICCS, KNOW, PSNC, SZE, and USTUTT. It also encompasses perspectives of the Flee Workshops and the online offer.

ECMWF will provide in Y2 and Y3 training on using and understanding ECMWF data, as well as on the use of ECMWF custom RestAPIs [21] for obtaining data (the Copernicus CDS API [28] and the Weather and Climate Data API – WCDA [29]). This training will be provided to HiDALGO partners working on the Migration (BUL) and UAP pilots (SZE), as well as to partners working on integrating HiDALGO workflows within the HiDALGO framework (PSNC, USTUTT, and ATOS). Training on using WCDA for data access will be a part of training activities for users of the European Weather Cloud [30].

ECMWF will also develop a set of Jupyter Notebook trainings on obtaining and manipulating weather, climate, hydrological, and air quality data, aimed at users beyond the meteorological community. These data, for the specific period or area of interest, can be also provided by ECMWF to support the HiDALGO consortium in training events or workshops. If need be, ECMWF will contribute with dedicated training sessions.

ICCS will integrate HiDALGO-related HPDA techniques and results into under- and postgraduate courses offered by the National Technical University of Athens (Greece):

- Parallel Processing Systems (9/2020–1/2021 and 9/2021–1/2022)
- Distributed Systems (9/2020–1/2021 and 9/2021–1/2022)
- Large scale data management (3/2020–6/2020 and 3/2021–6/2021)
- Parallel Computations and AI (3/2020–6/2020 and 3/2021–6/2021)

KNOW currently offers an HPDA training curriculum, which includes the following teaching activities: Big Data and ML, Knowledge Discovery and Data Mining, Statistics and Data Mining, Introduction to Data Science and AI. Furthermore, the tutorial "AI Essentials Training" (presentations and hands-on exercises) mainly targets team and project leads, innovation managers, process owners, and system administrators. In Y3, selected events will be adapted to the specific training needs of the consortium in terms of AI.

PSNC is going to participate actively in all project undertakings, such as conferences, summer schools, and training. Moreover, the infrastructure described in Section 4.2 will be exploited to offer relevant knowledge in the use of HPC systems. PSNC will prepare material dedicated to the online training system (Moodle, Section 5) as well as live training to users who are able to participate in person. All activities are conducted free of charge by experienced local HPC staff as well as by invited lecturers. Training topics include:

- Introduction to the purposes of HPC computing and overview on basic principles: Main components of HPC systems, working with queue systems, different kinds of parallelisms on a supercomputer.

- Topological optimization and parametric optimization.
- Development of modern authentication and authorisation mechanisms.

SZE plans to develop and launch a full course on UAP for policy makers in Y2. The course will include a section on the HiDALGO infrastructure and will consist of lectures on basic concepts of modelling and simulation, HPC, and data analytics. Then, the course will continue with lectures on the air pollution challenge, the air quality directives of the European Commission, the UAP workflow, and a hands-on tutorial on using UAP.

Plans for Y2 and Y3 foresee the enrichment of (at least) three USTUTT courses [20] with a dedicated session elaborated and presented by HiDALGO personnel, namely:

- Parallel Programming Workshop (MPI, OpenMP and Advanced Topics): Inclusion of the MPI Python API in addition to Fortran and C/C++, to bring the participants closer to some packages extensively used within HiDALGO (mpi4py, dask-distributed).
- Iterative Linear Solvers and Parallelization: A lecture and a hands-on session on computing eigenvalue histograms (eigenvalue estimators) of large-scale sparse symmetric matrices with PETSc, e.g. applied to large scale networks.
- CFD with OpenFOAM®: Undergoing plans include a demonstration of the UAP case study, in particular the results obtained with OpenFOAM®. A brief presentation of the UAP had already been integrated in the course in March 2020, with a pointer to the UAP spokespersons for further information and details (course cancelled).
- ML, DL, and GPU programming using OpenACC [27]: A session by HiDALGO personnel on Data Compression of numerical data sets with the BigWhoop library is going to be included in the DL course offered by NVIDIA in July 2020. A second ML course is in preparation and scheduled for Y2.

Moreover, the First Flee Workshop in Adama, Ethiopia (Section 4.1), was received very favourably by the local community. As a result, this workshop is likely to be organised once more in Y2 or Y3. In fact, HiDALGO received two such invitations already during Q4/2019 and Q1/2020, but both workshops had to be cancelled at short notice due to local political instabilities. Because of this major uncertainty in the workshop organisation, it is not yet known how many workshops will be organised in the future, either in person or through remote teaching.

Concerning the online offer, the presentations and trainings provided at the multi-day workshops (Section 3.3) could serve as the basis for a Massive Open Online Course (MOOC). However, a decision about this will be made according to the availability of the workshop speakers.

To enrich the webinar series mentioned in Section 4.1, USTUTT plans to host further webinars on the HiDALGO Portal modules, for example:

- HiDALGO Portal Webinar: A public webinar to promote several services available in the HiDALGO Portal (CKAN, Cloudify, Moodle, Matchmaking, Visualization, Support System, Interactive Notebook, see D5.3 [31]).
- CKAN and Cloudify Webinar: A webinar within the consortium with the goal of harmonizing the tools inside the project. Cloudify is a meta-scheduler to define and orchestrate the application workflow. CKAN is a data management tool to store input and output data of simulations and provide data as a service to integrate all the data.

5 Best practices

HiDALGO events performed so far and mentioned in Sections 3.1 and 4.1 allowed us to draw the conclusions below regarding best practices for WP7 activities and in particular for training.

According to the HiPEAC experience (Section 3.1), it has been deemed advisable to disseminate future events with a dedicated website and with live-tweets during the session. This time, the event has been advertised through the partners' networks, Twitter, and the homepage.

Moreover, this first training opportunity on the UAP case study allowed the project to improve existing and develop new teaching material. Feedback on the material developed for HiPEAC has been positive: HiPEAC attendees understood the goals of the HiDALGO UAP pilot and acknowledged the used technology. Based on the material of the HiPEAC course, completed with new developments especially in the preprocessing phase, SZE has been constructing a complete educational course to be offered in Y2 (see Section 4.3).

On the technical side, an infrastructure for the hands-on session at HiPEAC has been designed, to allow for anonymous access by the users and execution of jobs without any delay (compared to HPC systems). Furthermore, a dedicated synospsis has been prepared to conduct and explain all the necessary commands as a "workflow implementation". More advanced training could be offered based on these tools, which will be part of the Moodle service described below.

The tutorials in the First Flee Workshop (Section 4.1) were well-received, and the format of the three-day event was an effective way to raise awareness and develop skills among the participants. Meanwhile, the tutorial on prototyping self-defined simulations was judged relatively difficult compared to the other tutorials, so that it has henceforth been enriched with additional explanations.

Moreover, in order to technically improve the teaching and learning experience, a Moodle portal has been set up [42]. Moodle is an online platform to conduct courses and share training material, which allows access to training material in a structured and interactive way. A Course Instructor is responsible for organising the course structure, for managing the student enrolment, and for maintaining the corresponding course modules in the market. The courses can be structured with various features and best practices to ensure both a continuous improvement in content and structure, and the courses' appeal for stakeholders and students. The following items present a summary of generic best practices in structuring courses in order to create an impact on the market:

- Creating sub-modules for the course: One sub-module covers one specific week for optimal structure and timeline
- Introduction at the beginning of each sub-module

- Summary at the end of each sub-module
- Definition of participants' prerequisites and course outcome
- Sharing slides, graphical artefacts, references (books, scholar) and linked videos to give a complete set of documents for enrolled students and stakeholders
- Setting up quizzes and feedback forms for students to evaluate the content
- Analysing the course metrics (successfully completed, dropped students)
- Ensuring the course outcome has been achieved and identifying possible improvements in the course content and structure
- Improving the course content based on the students' feedback and course metrics
- Collecting all common definitions throughout the project (internal)

Video-recording and other complementary material will be made available online, by means of a fundamental cooperation with other tasks, such as T7.4 (Dissemination and Communication), and the technical tools made available by WP5 (Centre Implementation and Operation). All the webinar recordings have already been archived in GoToWebinar [41] and their link shared with the partners. An overview of training events and other HiDALGO activities can be found in the bi-annual newsletter, while scientific publications are made public through dedicated channels, e.g. ResearchGate.

Another important point is inclusivity: Educational tasks should be accessible to attendees with limited means (students, people from developing countries, etc.). Exemplary in this sense has been the First Flee Workshop (Section 4.1), where the training has been provided on site, with no fees for the participants. Also, compared to in-person training, webinars are an inclusive training practice, which drastically reduces participation fees and travel costs. Thus, when organising the next training activities together with T7.2 (e.g., the Y3 multi-day workshop), a travel cost subsidy and a "fair" fee system will be thought of, considering status and origin of the participants. For reference, at USTUTT courses, fees already vary according to student status, and academia or industry affiliation.

6 KPIs and further initiatives

Functional and quality-related KPIs for T7.3 have been defined at a preliminary stage of this task. The quality-related KPIs refer to the ACSI Score (American Customer Satisfaction Index), a set of three questions assessing satisfaction, performance, and expectation levels, each on a different 10-point scale (see e.g. Section 2.3 of [34]). In Table 1, the targets for Y2 and Y3 are reported, together with the corresponding achievements (from the start of the activities until M15).

KPI number	Description	Target Y2	Target Y3	Score at M15
7.3.1	Number of training / workshop activities done	≥1	≥2	4
7.3.2	Number of participants trained / educated	≥10	≥20	79
7.3.3	Satisfaction degree of participants	≥70%	≥85%	Not measured
7.3.4	Percentage of participants outside the consortium	≥60%	≥80%	Not measured

Table 1. KPIs for T7.3

KPIs 7.3.1 and 7.3.2 have already far exceeded the targets set for Y2 and Y3, which is a remarkable result for HiDALGO. On the other hand, KPIs 7.3.3 and 7.3.4 could not be measured, since they have been defined after the start of T7.3 within WP7 meetings. Thus, the necessary tools to measure them could not be finalized before the events considered in Table 1 took place. In particular, a questionnaire (online or on paper, with a content dependent on the nature of the event) should be distributed to the participants at the conclusion of each event (KPI 7.3.3). The information in KPI 7.3.4 should be directly requested at registration stage. Nevertheless, it can be stated that all the participants (60) at the First Flee Workshop (Section 4.1) and at the HiPEAC event (Section 3.1) do not belong to the consortium, while those who attended the two webinars (19) (Section 4.1) belong to either HiDALGO or EXCELLERAT [13].

For future reflection, we considered a few other useful indicators. First of all, we extracted the following information from the participants at the training activities:

- Stakeholder: Two industrial participations at the webinars were recorded (there is unfortunately no record available for the HiPEAC event).
- Country of origin (EU vs. non-EU): A strong predominance of EU participants has been observed in three of the four activities evaluated in Table 1, while all the participants at the First Flee Workshop have local origin (Ethiopia).

Another key issue is the gender ratio of trainers and trainees. By monitoring it, HiDALGO activities are fostered to promote gender-balance and inclusivity. This is a common effort with other CoEs involving training, e.g. EXCELLERAT (D5.1 [34] and D5.3 [35]). So far, concerning the four events in Table 1, about 1/6 of both the participants and the HiDALGO personnel involved in delivering training is female. Concrete initiatives to improve these figures include the involvement of HiDALGO in WomenInTech [39] and Women in HPC [40]. Moreover, communication channels dedicated to gender equality (e.g., gender equality mailing lists at universities) will be further identified and used to spread the messages. Once established, the Coordinartion and Support Action (CSA) for the National Competence Centres will be taken in the loop to circulate HiDALGO activities to the diverse European countries and stakeholders.

7 Conclusions

While it is encouraging that training activities have begun before the actual start of the task, it is essential to keep up the pace in offering training activities and organising (training) events. In fact, these require a thorough and early preparation, a constant contact with the technical developers within the CoE, and a monitoring of the advancement of the pilots.

While bringing together GC and HPC/HPDA communities is a realistic goal of "innovation workshops", providing training targeted at fulfilling specific needs will be a more demanding achievement. Nevertheless, the existing training and support offer, as well as the ongoing contact with the stakeholders are promising steps in this direction.

The results of this document will be used as input in the upcoming deliverable D7.4 (Annual Report on Community Building, Event Management, Collaboration and Training), followed by D7.5 and D7.6. The content of training activities will be an important part of the Portal, therefore T7.3 will also contribute to D5.6 (Second HIDALGO Portal Release and System Operation Report), followed by D5.7.

References

[1] HiPEAC 2020 (Bologna, Italy, 20-22 January 2020), [Online] https://www.hipeac.net/2020/bologna/#/ [Accessed 2020-03-05].

[2] [Online] https://www.hipeac.net/2020/bologna/#/schedule/sessions/7722/ [Accessed 2020-03-05].

[3] Open Science Week at the Adama Science and technology University, [Online] https://agenda.ethernet.edu.et/event/57/program [Accessed 2020-03-05].

[4] International Symposium: Urban Systems – Global Challenges – Digital Tools, (Stuttgart, Germany, 28-29 May 2019), [Online] https://www.hlrs.de/events/archive/detail-view/2019-05-28-ugd/ [Accessed 2020-03-05].

[5] Forum TERATEC 2020 (Palaiseau, France, 16-17 May 2020), [Online] http://www.teratec.eu/forum/index.html [Accessed 2020-03-05].

[6] ICCS2020 (Amsterdam, The Netherlands, 3-5 June 2020), [Online] https://www.iccs-meeting.org/iccs2020/ [Accessed 2020-03-05].

[7] Multiscale Modelling and Simulation Workshop, 17th edition (Amsterdam, The Netherlands, 3-5 June 2020), [Online] https://mms.computationalscience.nl/ [Accessed 2020-03-05].

[8] ISC High Performance (Frankfurt am Main, Germany, 21-25 June 2020, Germany), [Online] https://www.isc-hpc.com/ [Accessed 2020-03-05].

[9] Social Simulation Conference (Milan, Italy, 14-18 September 2020), [Online] http://ssc2020.behavelab.org/ [Accessed 2020-03-05].

[10] International Conference on Principles and Practice of Multi-Agent Systems - PRIMA 2020 (Nagoya, Japan, 27-30 September 2020), [Online] http://uchiya.web.nitech.ac.jp/prima2020/ [Accessed 2020-03-05].

[11] HiDALGO CoE website, [Online] https://hidalgo-project.eu [Accessed 2020-03-05].

[12] FocusCoE CSA website, [Online] https://www.focus-coe.eu/ [Accessed 2020-03-05].

[13] EXCELLERAT CoE website, [Online] https://www.excellerat.eu/ [Accessed 2020-03-05].

[14] EoCoE website, [Online] https://www.eocoe.eu/ [Accessed 2020-03-05].

[15]ESiWACE CoE website, [Online] https://www.esiwace.eu/ [Accessed 2020-03-05].

[16]PRACE (Partnership for Advanced Computing in Europe) Events Portal, [Online] https://events.prace-ri.eu/ [Accessed 2020-03-05].

[17]ChEESE CoE website, [Online] https://cheese-coe.eu/ [Accessed 2020-03-05].

[18]OpenProject Webinar, [Online] https://www.excellerat.eu/wp/events/?event_id1=104 [Accessed 2020-03-05].

[19]Best Practice Guide for Git & Jenkins Webinar, [Online] https://www.excellerat.eu/wp/events/?event_id1=96 [Accessed 2020-03-05].

[20]HLRS Upcoming Training Courses, [Online] https://www.hlrs.de/training/ [Accessed 2020-03-05].

[21]Fielding, Roy Thomas (2000). "Chapter 5: Representational State Transfer (REST)". Architectural Styles and the Design of Network-based Software Architectures (Ph.D. thesis). University of California, Irvine.

[22]Sergiy Gogolenko, Derek Groen, Diana Suleimenova, Imran Mahmood, Marcin Lawenda, F. Javier Nieto De Santos, John Hanley, Milana Vučković, Mark Kröll, Bernhard Geiger, Robert Elsässer, and Dennis Hoppe. “Towards Accurate Simulation of Global Challenges on Data Centers Infrastructures via Coupling of Models and Data Sources” / Preprint.

[23]Apache Spark™, [Online] https://spark.apache.org/ [Accessed 2020-03-25].

[24]Apache Kafka®, [Online] https://kafka.apache.org/ [Accessed 2020-03-25].

[25]NEK5000, [Online] https://nek5000.mcs.anl.gov/ [Accessed 2020-03-05].

[26]OpenFOAM®, [Online] https://www.openfoam.com/ [Accessed 2020-03-05].

[27]OpenACC, [Online] https://www.openacc.org/ [Accessed 2020-03-24].

[28]Copernicus CDS API, [Online] https://cds.climate.copernicus.eu/api-how-to/ [Accessed 2020-03-05].

[29]HiDALGO D4.2 – Implementation Report of the Pilot Applications. Horváth, Zoltán 2019.

[30]HiDALGO D5.1 – HiDALGO System Environment. Kovács, Ákos 2019.

[31]HiDALGO D5.3 – First HIDALGO Portal Release and System Operation Report. Nieto de Santos, Francisco Javier 2019.

[32]HiDALGO D6.4 – Initial Report on Requirements, Components and Workflow Integration. Tsoumakos, Dimitrios 2019.

[33]HiDALGO D7.2 – Annual Report on Community Building, Event Management, and Collaboration. Elsässer, Robert 2019.

[34]EXCELLERAT D5.1 – Initial Assessment of Training Needs and Services Building Plan. Arlandini, Claudio Domenico 2019.

[35]EXCELLERAT D5.3 – Services, Training and Portal Report. Arlandini, Claudio Domenico 2020.

[36]FocusCoE D4.3 – Report on the outcomes of the comprehensive European stakeholder workshop on HPC training and skills development needs. Wong, Simon 2019.

[37]Gesellschaft für internationale Zussammenarbeit GmbH, [Online] www.giz.de [Accessed 2020-03-05].

[38]International Organization for Migration, [Online] https://www.iom.int [Accessed 2020-03-05].

[39]Women in Tech, [Online] https://womenintech.fi/ [Accessed 2020-03-05].

[40]WHPC, [Online] https://womeninhpc.org/ [Accessed 2020-03-05].

[41]GoToWebinar, [Online] https://www.gotomeeting.com/webinar [Accessed 2020-03-05].

[42]Moodle, [Online] https://docs.moodle.org/38/en/Managing_a_Moodle_course [Accessed 2020-03-05].

Annexes

Date	Name of the event	Location	Type	Section
05/2019	International Symposium: Urban Systems – Global Challenges – Digital Tools	Stuttgart, Germany	Innovation Workshop	3.3
07/2019	First Flee Workshop	Adama, Ethiopia	Educational Task	4.1
08/2019	OpenProject Webinar (USTUTT)	Web	Webinar	4.1
10/2019	Best Practice Guide for Git & Jenkins (USTUTT)	Web	Webinar	4.1
01/2020	HiPEAC Conference	Bologna, Italy	Innovation Workshop	3.1
05/2020	Forum TERATEC 2020	Palaiseau, France	Fair	3.3
06/2020	ISC High Performance 2020	Frankfurt am Main, Germany	Innovation Workshop	3.3
06/2020 (or 2021)	Multiscale Modelling and Simulation Workshop	Amsterdam, The Netherlands	Innovation Workshop	3.3
07/2020, Q3/2020, 2021	Deep Learning and GPU programming using OpenACC	Stuttgart, Germany	Educational Task	4.3
09/2020 (or 2021)	International Conference on Principles and Practice of Multi-Agent Systems - PRIMA 2020	Nagoya, Japan	Innovation Workshop	3.3
09/2020	Global Challenges: Risk and Complexity	Stuttgart, Germany	Innovation Workshop	3.3
09/2020	Social Simulation Conference	Milan, Italy	Innovation Workshop	3.3
09/2020 to 01/2022	ICCS HPDA Training in courses at the National Technical University of Athens	Athens, Greece	Educational Task	4.3
2020/2021	ECMWF HiDALGO-specific Training	tbe	Educational Task	4.3

2020/2021	PSNC Training with PSNC Virtual Machines	Poznań, Poland	Educational Task	4.3
2020/2021	SZE Training on UAP	tbe	Educational Task	4.3
2020/2021	USTUTT HiDALGO-specific Training	Stuttgart, Germany	Educational Task	4.3
2020/2021	Further Flee Workshops	Adama, Ethiopia / Web	Educational Task	4.3
2020/2021	Webinars on HiDALGO Portal Modules (USTUTT)	Web	Webinar	4.3
Q1/2021	CoEs joint Winter School	Stuttgart, Germany / tbe	Innovation Workshop	3.3
06/2021	ISC High Performance 2021	Frankfurt am Main, Germany	Innovation Workshop	3.3
2021	KNOW HiDALGO-specific AI Training	tbe	Educational Task	4.3
2021	HiDALGO-initiated Workshop	tbe	Innovation Workshop	3.3

Table 2. HiDALGO Training Curriculum