D7.3 Training Concept

Document Identification			
Status	Final	**Due Date**	31/03/2020
Version	2.0	**Submission Date**	21/01/2021

Related WP	WP7	**Document Reference**	D7.3
Related Deliverable(s)	D4.2, D5.1, D5.3, D7.2, D7.4	**Dissemination Level (*)**	PU
Lead Participant	USTUTT	**Lead Author**	Lorenzo Zanon
Contributors	BUL, DIA, ECMWF, ICCS, KNOW, PLUS, PSNC, SZE	**Reviewers**	Alexandra Gens (DIA)
			John Hanley, Claudio Iacopino (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).
1.1	04/12/2020	L. Zanon (USTUTT)	Revisited document after request for revision 30/10/2020 for QA.
1.2	11/12/2020	L. Zanon (USTUTT)	Version after internal review.
2.0	14/01/2021	L. Zanon (USTUTT)	Final version after Quality manager and Project Coordinator 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
Deliverable leader	Lorenzo Zanon (USTUTT)	14/01/2021
Quality manager	Marcin Lawenda (PSNC)	19/01/2021
Project Coordinator	Francisco Javier Nieto (ATOS)	21/01/2021

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
ABMS	Agent-based modelling and simulation
AI	Artificial Intelligence
API	Application Programming Interface

Abbreviation / acronym	Description
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
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 effort conducted by HiDALGO Task 7.3 (T7.3) in bringing together the Global Challenges (GC) and the High Performance Computing (HPC) - High Performance Data Analysis (HPDA) communities by providing a two-track training curriculum. This comprises on the one hand events for HPC technology leaders, on the other hand events for scientists and analysts who tackle Global Challenges through the support of HPC and HPDA. The quality of the HiDALGO curriculum depends on the content delivered, as well as on both didactic and technical best practices in training. At the same time, training events are opportunities to exchange among the different communities, and to collect feedback on the technical pilots of the project. In summary, this document describes the HiDALGO training concept in terms of training requirements, current and planned events, best practices, and collected feedback.

HPC and HPDA training requirements in GC arise from the development of the use cases, while 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). Both kinds of requirements represent one of the motivations for HPC/HPDA 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.

One efficient source of training is 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 online training tools such as web-seminars 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.

Finally, the results of this document will be the basis for the future WP7 deliverables D7.4 and D7.5. Moreover, the content of training activities will be an important part of the HiDALGO portal, therefore T7.3 will also contribute to the portal releases, subject of D5.6 and D5.7.

1 Introduction

1.1 Update to the first submission

This new version of deliverable D7.3 addresses the issues of the "Request for revision of deliverable submission" received on 30/10/2020. Specifically, the HiDALGO two-track training curriculum has been characterised in Section 2 in terms of requirements (Sections 2.1 and 2.2), features (Section 2.3), and topics (Table 11). The conducted and planned training activities are listed in Sections 4.1 and 4.2 with tables specifying the corresponding curriculum track, the addressed requirements, and the learned skills in each activity. In particular, the learned skills of the training offered by ICCS and PSNC (Table 8) have been enriched with a learning outcome beyond HPC basics, SZE training (Table 9) has among the objectives the replication of the UAP pilot, pySpark competences have been located at ICCS and USTUTT (Section 2.1) and will be reflected in future training activities. In Section 6, one additional KPI has been reported, and a further improvement of the KPI scores is referenced to deliverable D7.4.

Sections 2, 3, and 4 have a different structure compared to the first submission to enhance the document's readability. At the same time, the content of Sections 3 and 5 (Innovation Workshops and Best Practices) has not significantly changed, since it was not requested in the revision letter.

Since this deliverable was originally written in February and March 2020, the consequences of the COVID-19 pandemic on planned training events are only marginally touched upon. Updated information can be found in the subsequent deliverable D7.4.

1.2 Purpose of the document

This deliverable defines the HiDALGO training concept as a two-track curriculum, comprising both events for HPC technology leaders and for scientists and analysts who tackle Global Challenges through the support of HPC and HPDA. The common aim of such events is to raise mutual awareness amongst the GC and the HPC/HPDA communities. Subsequently, this deliverable analyses the current state-of-the-art of HiDALGO training activities and innovation workshops, and defines the necessary roadmap for future events. Other equally important aspects of T7.3 that are described in this document include the collection of feedback on the pilots from the stakeholders and the academic community, the interaction with other CoEs, and the identification of best practices in education and training.

1.3 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 [27] from the point of view of training, and will be updated by the training sections of D7.4 [28] and D7.5 [29]. 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 setup of technical tools for training best practices depends on the activity of WP5, dedicated to the establishment of the HiDALGO Portal (see D5.3 [25]). 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.

1.4 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 the HiDALGO two-track training curriculum for both HPC/HPDA experts and GC scientists.
- Section 3 details the joint effort of T7.2 and T7.3 to enable HiDALGO to play a proactive part at “innovation workshops” to promote the interaction between HPC/HPDA and GC communities.
- 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 The HiDALGO training concept

One of the main goals of T7.3 is to create a two-track training curriculum. On the one hand, a curriculum for HPC technology leaders must detail the Global Challenges needs, on the other hand a curriculum for GC scientists and analysts must focus on HPC and HPDA modelling. In the end, the common goal of the consortium is to both acquire and provide tools to tackle GC from an intertwined HPC and HPDA perspective.

This section describes the requirements collected from both the pilots (Section 2.1) and the stakeholders and participants (Section 2.2), and gathers the features of both tracks of the HiDALGO training curriculum (Section 2.3).

2.1 HPC and HPDA requirements from the Pilots

In the frame of the training curriculum track aimed at GC scientist, this section details HPC and HPDA requirements that emerged when simulating GC in the development of the HiDALGO pilots. In case a training has been offered or is planned, the training provider is indicated for each item. In Table 1 to Table 10 in Section 4 it is indicated whether any of these requirements is tackled by the described event.

GC simulations have given rise to a number of new requirements of HPC environments, which have been elaborated within the HiDALGO use cases as described in detail in [22]. A summary of these requirements follows, with a shortcut (e.g., [HPC_R1]) associated to each of them and reported in the training activities of Section 4:

- *Support for agent-based modelling paradigm* [HPC_R1]: In contrast to traditional HPC applications from engineering and science domains, GC applications often contain parts related to social simulations. Agent-based modelling 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.).
  Training provider: BUL (training on Flee in particular, see Table 1).
- *Coupling with both static and streaming external data sources (integration of streaming data into HPC environments)* [HPC_R2]: 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 are coupled with those of the

Migration and SNA pilots. Apache Kafka® [24] in particular is a distributed streaming platform that stores streams of records in a fault-tolerant durable way and processes them as they occur.
Training provider: PSNC, after further development planned in Y2 and Q1/Y3.

- *Coupling simulations across data-centres* [HPC_R3]: 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, weather forecast simulations and meteorological data (ECMWF) are coupled with CFD and ABMS simulations (Migration and UAP pilots) on the resources of both USTUTT and PSNC.
  Training provider: SZE supported by ECMWF.

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 [26]):

- *Support for high-performance, general-purpose analytics runtime(s)* [HPDA_R1]: Pilots have registered and to-date required method implementations over huge numerical data sources (UAP: standard matrix operations, aggregations, etc.), graph-analytics operations (SNA: link detection, clustering, etc.), and streaming data crunching and coupling (for all the three pilots, using sensor traffic and weather data).
  Training provider: KNOW (graph-analytics), ICCS.
- *Support for and ability of coupling the aforementioned runtime(s) with high-performance datastores or distributed file systems* [HPDA_R2] such as HDFS, Lustre, CKAN, relational/main memory databases, Apache Avro/Parquet, etc. (for all the three pilots).
  Training provider: USTUTT and PSNC (web-seminars planned).

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 Python and R* [TOOLS_R1]: HiDALGO pilots extensively use the Python and R ecosystems for data analytics and visualisation 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 SNA pilots.
  Training providers: USTUTT, ECMWF, KNOW.
- *Support for Apache Spark™ [33]* [TOOLS_R2], 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 utilised under R, Java and, especially from the data scientists' point of view, Python (through pySpark and Jupyter Notebooks).

Training provider: USTUTT, ICCS.

Finally, training on, e.g., scheduling or version control tools is relevant for developers' as well as for non-technical activities, and will also be provided by HiDALGO [TOOLS_R3].

2.2 Stakeholders' needs and training survey

Stakeholders' needs in terms of both HPC and HPDA knowledge will be detailed through a broader 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) through an ad hoc survey, to be both added on the CoE webpage [10] and circulated at innovation workshops (Section 3). 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 the project.

In the stakeholders' survey, a questionnaire about training will be formulated as:

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 parallelisation techniques.*
- ☐ *Code Optimisation 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]*

The same questionnaire to inquire on training needs and wishes will be included in the quality assessment to collect participants' feedback on each project's event, when possible. HiDALGO contributions to external activities might have to rely on different surveys, only partially or not at all customisable.

Results of the combined stakeholders' and participants' training needs, and how they will be implemented in the HiDALGO two-track training curriculum, will be presented in the next deliverable D7.4 [28]. Training already addressing topics in the list above will be associated to the shortcut [QUEST] in the training activities tables of Section 4.

2.3 The two-track training curriculum

This section provides for both training tracks a brief description and a list of prototypes to reach the respective training objective, based on the training experience so far and on future plans.

Track "Training on GC": It can be carried out with different levels of detail considering different audiences: HPC developers, but also HPC users or young scientists. This kind of training refers in general to the pilots' scenarios, where Migration and UAP are deemed the most mature pilots to provide training (see e.g., the Flee Workshop in Section 4.1 and HiPEAC 2020 in Section 3.1). Considering the activities HiDALGO has been organising, attending or planning, GC training for HPC (future) experts can happen in those forms:

- Training activities in the established training program of partners, with integrations of HiDALGO-specific contributions. Here, demonstrations of the pilots or lectures on methods used to tackle a specific GC can be offered. The consortium partners provide training for different communities (students, researchers and post-graduates, industry) with different levels of HPC knowledge.
- Full multi-day courses aimed at reproducing a specific pilot, where details about the tackled GC can be given, and the developed tools demonstrated with hands-on sessions.
- Recorded videos which include an overview of a GC, and separate hands-on sessions on the utilised tools, to also increase awareness of the applicability of the HiDALGO achievements. For this, remote access to the platform where the software has been deployed must be set up and granted to the participants.
- Interactive web-seminars where the project mission and one or more pilots are illustrated (with the HiDALGO Portal as a starting point, when ready), with extensive space left for questions and discussions.
- Tutorials for specific tools on data processing workflows and multiscale coupling, which are recurrent scenarios within GC simulation (see Section 2.1).
- Presentations, round tables, panels, and virtual booths at innovation workshops, where HPC scientists are able to approach HiDALGO personnel and establish collaborations on GC themes. Since innovation workshops have further goals beyond training (collection of feedback, dissemination and outreach, etc.), they are mostly monitored by T7.2. Those deemed relevant for training are listed in Section 3.

Track "HPC and HPDA training": This kind of training for GC scientists can take advantage of HiDALGO partners' expertise in terms of HPC and HPDA (Big Data, Machine/Deep Learning, Artificial Intelligence, etc.). The partners are prompted to contribute HiDALGO-specific training sessions on new tools and developed techniques. Towards the end of the project, the consortium envisages to offer complementary training resulting from the optimised codes of the pilots.

Specific topics can be narrowed down in two ways. First, the stakeholders' and training participants' survey (Section 2.2) is an indication of the demand for training on general HPC and AI topics. On the other hand, the project strives to provide a training curriculum tailored to the pilots' technical needs. Therefore, the gaps in HPC and HPDA knowledge of experts working on GC have been identified and listed in Section 2.1. Taking all this into account, HPC and HPDA training can happen in those forms:

- Training activities in the established training programs of partners, integrated with training on tools specifically used within HiDALGO, when possible corresponding to the needs listed in Section 2.1.
- Web-seminars or workshops on topics directly emerging from the listed requirements, e.g., on tools for high-performance data-storage (such as CKAN) or agent-based modelling and simulation (such as Flee).
- HPC face-to-face or online training and training material, on platforms where the pilots' software has been deployed (e.g., the PSNC platform). Remote access and technical support should be provided.
- Advanced HPC training sessions in more general HPC courses, to focus on benchmarking procedures, optimisation, and coupling (also applied to the pilots).
- Training on HPDA with direct application to techniques used within the pilots, or training on AI components developed in the project.
- Specific training on obtaining, using and understanding weather data, which are relevant in many GC simulations (e.g., Migration and UAP). ECMWF is the natural partner for this endeavour. Moreover, ECMWF develops training on Jupyter Notebooks and on cloud computing.

Even though the two tracks should complement each other, overlapping is not excluded and is even beneficial for the exchange between different communities. For instance, whereas debates on GC or HPC tutorials clearly address either training track, training on specific HPC/HPDA tools for GC can be ascribed to both tracks. The same holds for training on, e.g., software for scheduling or version control, which is certainly useful to most members of HiDALGO and other CoEs. Moreover, the GC and HPC/HPDA background is not uniform among participants in training activities, who could thus have an interest in events of both tracks.

Together with the innovation workshops (Section 3), the education activities described in Section 4 are the fundamental part of the two-track training curriculum: An overview of all performed and planned activities building up the HiDALGO curriculum can be found in Table 13 (Annexes). Section 4 is wrapped up by Table 11, where the two tracks are mapped to the corresponding technical topics, covered by planned and performed activities, as well as those included in the technical requirements in Section 2.1.

The quality of the training curriculum is finally assessed and 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". From the point of view of training, they have the multi-fold purpose of collecting technical feedback on HiDALGO's activities from partners and stakeholders, 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 [27]): Workshops and sessions at conferences are planned with the shared goal of driving mutual GC and HPC/HPDA awareness, disseminating results and methods, and educating the participants. HiDALGO strives to organise some of these events jointly with other projects, e.g., with EoCoE-II [13] (first contact currently planned), ESiWACE-2 [14] (established contact regarding the use cases), or EXCELLERAT [12] (see Sections 3.1, 3.2, 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.

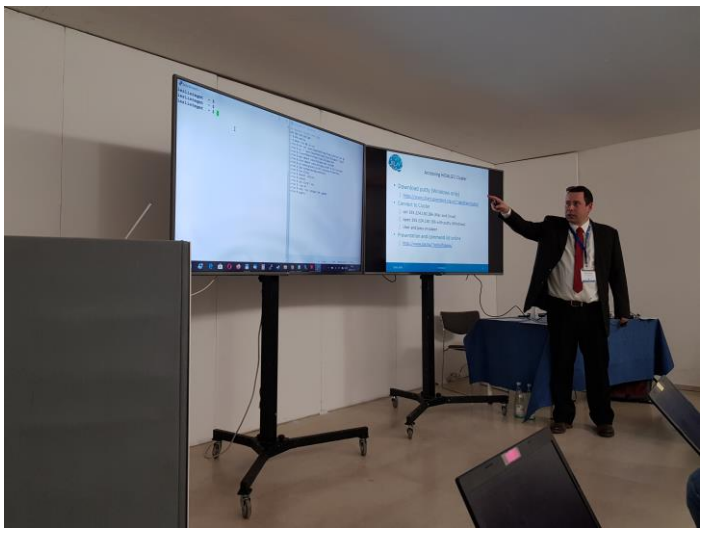

Figure 2. HiPEAC Workshop – UAP pilot presentation.

During the workshop, the audience showed interest in the SNA pilot, in particular in the Artificial Intelligence (AI) technologies used in there, leading to a discussion between scientists from the University of Pisa and members of the project about possible ways on how to cluster social networks.

There has also been a constructive interaction with the CoE EXCELLERAT [12], which is seeking new engineering applications to boost towards higher scalability and performance (see e.g., EXCELLERAT D5.1 [30] and D5.3 [31]). Within the UAP case study, collaboration plans with EXCELLERAT involve the evaluation of the HiDALGO CFD codes, the definition of optimal simulation parameters, the examination and improvement of the parallelisation, and optimisation at different levels (pre-processing, geometry preparation, and mesh generation). Even though speed and scalability should be first improved on OpenFOAM® (currently in use), HiDALGO could benefit from the EXCELLERAT flagship code NEK5000 [34], designed to solve incompressible Navier-Stokes flow simulations at low Mach number.

3.2 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 already been given in the 2019 edition of this event [3]. 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 [27]). 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 [4]). Feedback about the needs and requests arising from industry is already being collected by WP2 through the FocusCoE [11] network via an online survey.

HiDALGO will have a dedicated session at the workshop on Multiscale Modelling and Simulation [6] within the International Conference on Computational Science [5] (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 [7] (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) [9], and the Social Simulation Conference [8], 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 [15], 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).

A multi-day workshop is taken into account, possibly in cooperation with other CoEs or HPC centres, or as an initiative back-to-back to other established events. In a different fashion compared to conferences, this would be an occasion to raise awareness for GC and to build a

GC forum. This and similar events should encourage people interested in GC to network and exchange different perspectives towards creating a GC 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 by the ECM, will become a Y3 highlight. Additionally, a joint "winter school" in Y2 or early Y3 is in preparation with EXCELLERAT [12] and foresees the involvement of other CoEs (ChEESE [16], FocusCoE [11]). The (preliminary) focus lies on specific HPC topics (load balancing, in situ and remote visualisation, 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

Besides the innovation workshops of the previous Section, which have further goals beyond training (collection of feedback, dissemination and outreach, etc.), the HiDALGO two-track training curriculum consists mainly of activities focused on education. Among them, a number of HiDALGO-specific training sessions are integrated within the partners' established curricula of workshops and courses.

Such a knowledge transfer should happen in connection with the growing expertise collected through the case studies. During Y1 and Q1/2020 the Migration and UAP case studies have reached the maturity needed for training, while the SNA pilot still needs further technical development. Towards the end of the project, it is envisaged to offer complementary training resulting from the optimised codes.

Educational events are also an opportunity to cooperate with other CoEs to offer technical training, as it has also been framed by T7.2 (see D7.2 [27]).

4.1 Activities performed

Until M15, the training events in the tables below have taken place, and for each activity the curriculum track (Section 2) is indicated:

- Global Challenges needs for HPC technology leaders;
- HPC and HPDA modelling for GC scientists and analysts;
- Both: Event of interest for both communities, or training on tools useful to the project or to other CoEs.

It is also indicated whether the activity satisfies a particular training requirement (also indicated with the corresponding shortcut) among the list of HPC/HPDA requirements in GC applications in Section 2.1, and whether training material has been made available on the Moodle online platform (Section 5).

	1st Flee Workshop, Adama Science and Technology University
Curriculum Track	HPC and HPDA modelling
Type / format	Training workshop, Moodle material
Aim	Dissemination of Flee code, training, advertise HiDALGO work, try to get potential users for the case study, build up collaboration with the University of Adama.
Venue & date	Adama, Ethiopia, 16-18 July 2019.
Target group	Mostly academics (25 participants from the University of Adama)
Training requirement addressed	Agent-based modelling and simulation tools [HPC_R1].

	1st Flee Workshop, Adama Science and Technology University
Learned skills	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 and was carried out with the additional involvement from GIZ [18]. The agenda in short: Day 1: Python Software Carpentry Day 2: Introduction to Flee code, coupling and validation; tutorial "Design and Prototype your own Agent-based Simulation" (ABMS for addressing GC) Day 3: Hackathon around simulation building (modelling of (forced) migration) A tutorial on the usage of Flee has been made available on Moodle afterwards.
Impression	The tutorials were well-received, and the format of the three-day event was an effective way to raise awareness and develop skills among the participants. The participants stressed the importance of modelling "internally displaced people" as well as refugees, and also noted the wide range of modes of transport commonly adopted by displaced people in South Sudan. Positive feedback has been received from agricultural agencies as well as from several humanitarian organisations and stakeholders, including the UNHCR and the International Organization of Migration [18], which are currently being involved in the HiDALGO efforts to bolster the Migration case study. More training courses are planned in collaboration with the University of Adama, and HiDALGO received two such invitations already during Q4/2019 and Q1/2020. Due to political uncertainties and the COVID-19 pandemic, it is not yet known how many workshops will be organised in the future, either in person or through remote teaching.

Table 1. 1st Flee Workshop, Adama Science and Technology University.

	Introductory Webinar on OpenProject
Curriculum Track	Both
Type / format	Web-seminar
Aim	Explain features of OpenProject (an online project management tool), provide hands-on and best practices to use this tool within HiDALGO.
Venue & date	Online, 2 August 2019.
Target group	Mostly academia, Industry (HiDALGO and EXCELLERAT partners)
Training requirement addressed	Training on tools useful for managing the project [TOOLS_R3].
Learned skills	The goal of OpenProject is to manage all the technical activities within the WPs following the principles of Agile Software Development (ASD). Technical task assignments and project progress are tracked easily by OpenProject, with the effect of improving the clarity between multiple partner contributions and enhancing the cooperation among the partners.
Impression	The web-seminar was well received: Afterwards, all WP leaders created sub-projects within the HiDALGO main project to manage their activities by following the (ASD) principles (Figure 3 and Figure 4). In particular, Figure 4 depicts the WP5 backlog in 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. After this kick-off, OpenProject has actually played a minor role in the HiDALGO management. Nonetheless it is actively used by EXCELLERAT, which co-hosted the workshop.

Table 2. Introductory Webinar on OpenProject.

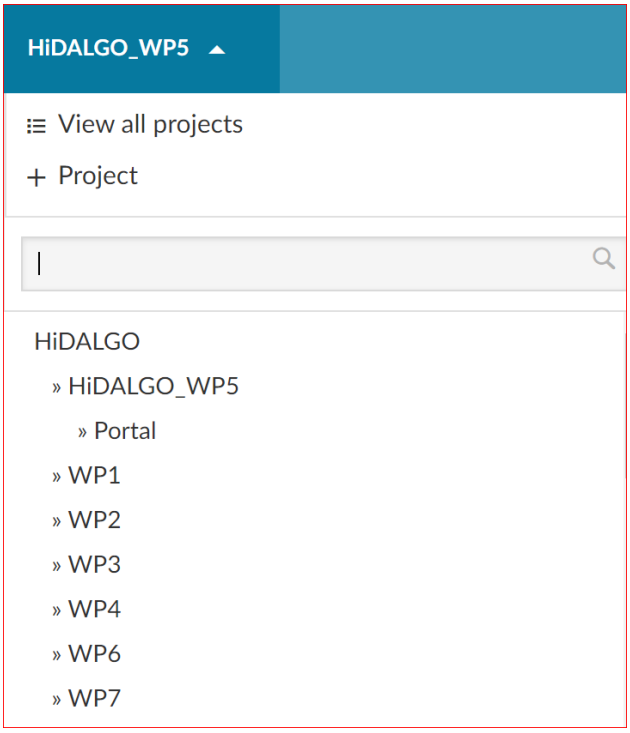

Figure 3. OpenProject: All the WP sub-projects are created and managed by the WP leaders.

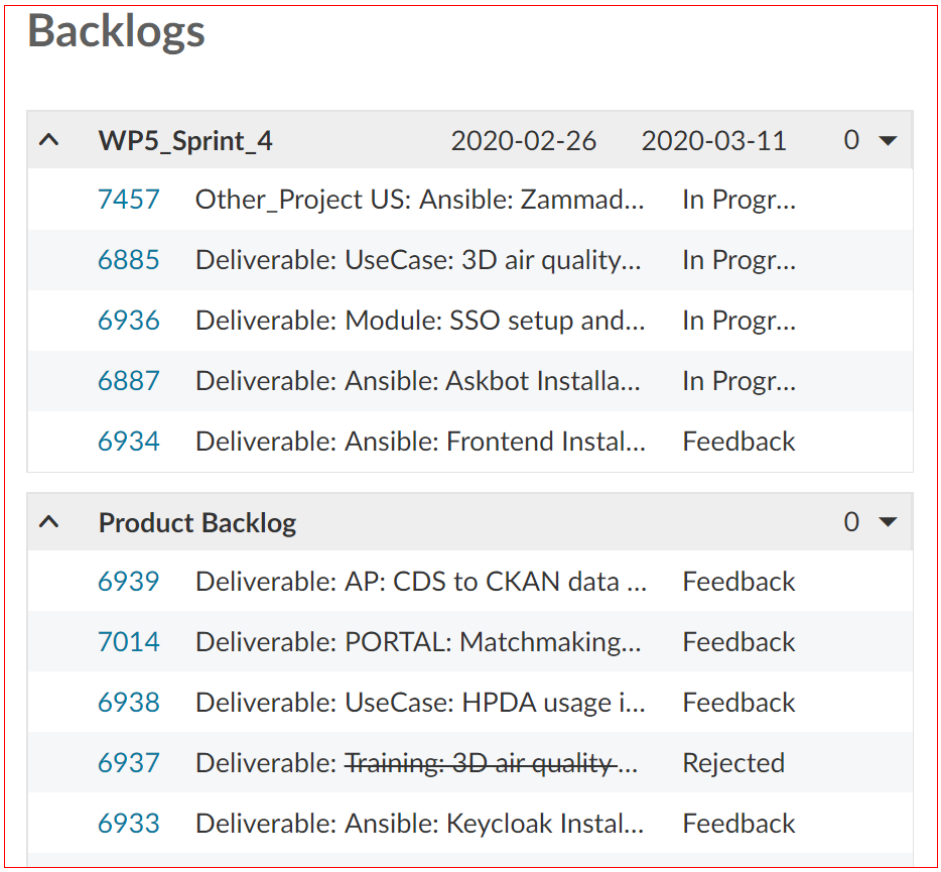

Figure 4. OpenProject: WP5 backlog with the current task information and the Product Backlog.

	Best Practice Guide for Git & Jenkins
Curriculum Track	Both
Type / format	Web-seminar
Aim	Explain the basic features of Git and Jenkins, provide hands-on and best practices to use these tools within HiDALGO.
Venue & date	Online, 31 October 2019.
Target group	Mostly academia, Industry (HiDALGO and EXCELLERAT partners).
Training requirement addressed	Training on tools for faster and collective developers' work [TOOLS_R3].
Learned skills	Jenkins and Git, hosted at USTUTT, are tools to safely manage the source code of the project and to automate software development. After the Git web-seminar, a central repository has been created for all the HiDALGO tools and technologies to manage the corresponding code (Figure 5).
Impression	HiDALGO developers are now actively using the Git repository to manage their source code. HiDALGO portal development follows Continuous Integration and Deployment (CI/CD), so the developers are actively using Git to automate the portal components deployment and testing. A questionnaire has been additionally circulated to investigate users' feedback about Git and Jenkins. 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.

Table 3. Best Practice Guide for Git & Jenkins.

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

4.2 Planned activities (roadmap)

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 integrate HiDALGO-specific training in their activities starting in Y2 or Y3.

A noticeable effort has been already conducted by PSNC in terms of infrastructure modernisation 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 have been deployed and configured on these machines. PSNC plans HPC training based on the Virtual Machines, see Table 8.

Upcoming activities in terms of web-seminars, videos, online activities, and of integrations to the partners' training programs through HiDALGO-specific initiatives are listed in the tables below. For more details on these activities, their outcome and future plans, the reader is referred to the subsequent deliverable D7.4 [28].

Web-seminars, videos, and online activities

	Cloudify Blueprint Preparation, CKAN API Usage
Curriculum Track	HPC and HPDA modelling
Type / format	Web-seminar
Aim	Mostly internal web-seminar to explain the tools Cloudify and CKAN.
Venue & date	Online, Q2/2020.
Target group	Participants from the consortium.
Training requirement addressed	High-performance datastores and distributed file systems [HPDA_R2].
Learned skills	This web-seminar plans an introduction to Cloudify (to help pilots to prepare Blueprint to define their workflow) and to CKAN (for managing their data). All written material will be made available through the Moodle afterwards.

Table 4. Cloudify Blueprint Preparation, CKAN API Usage.

Integrations to the partners' training programs through HiDALGO-specific initiatives

	Training by ECMWF on weather and climate data access and manipulation
Curriculum Track	Both
Type / format	Online and on-site training.
Aim	Provide training to the consortium members and externals.
Venue & date	Online or on-site, Y2 and Y3.
Target group	Academia and industry representatives, within and outside the consortium.
Training requirement addressed	Coupling of weather and climate data [HPC_R3], Support for Python [TOOLS_R1].
Learned skills	ECMWF will provide training on using and understanding ECMWF data, as well as on the use of the ECMWF REST APIs [21] for obtaining data (the Weather and Climate Data API – WCDA [23], the Copernicus Climate Data Store API [36], and Atmospheric Data Store API). This training will be provided to HiDALGO partners working on the Migration (BUL) and UAP (SZE) Pilots, 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 [24]. ECMWF will also develop a set of tutorial Jupyter Notebooks on obtaining, manipulating and visualising weather, climate, hydrological, and air quality data, aimed at users beyond the meteorological community. They will be shared with HiDALGO partners to help them work with the data, possibly on Moodle. ECMWF also plans to provide data for specific periods or areas of interest, to support the HiDALGO consortium in training events or workshops. If needed, ECMWF will contribute with dedicated training sessions.

Table 5. Training by ECMWF on weather and climate data access and manipulation.

	Training by ICCS on HPDA methods
Curriculum Track	HPC and HPDA modelling
Type / format	Online and on-site training
Aim	Provide training to the consortium members and externals.
Venue & date	School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Greece, autumn 2020 to the end of the project.
Target group	Academia representatives, mostly outside the consortium at the NTUA.
Training requirement addressed	Graph-analytics operations [HPDA_R1].

	Training by ICCS on HPDA methods
Learned skills	ICCS will integrate HiDALGO-related HPDA techniques and results into the following under- and postgraduate courses offered by the National Technical University of Athens: ▸ Parallel Processing Systems: This undergraduate course is offered every winter semester. ICCS will incorporate lectures on the HPC benchmarking procedure used in the HiDALGO project, focusing both on tools (such as ScoreP for profiling/tracing and Vampir for visualisation) and best practices. Complementary hands-on sessions will focus on code benchmarking and optimisation. ▸ Parallel Computations and AI: This postgraduate course is offered every spring semester. Similar to the "Parallel Processing Systems" undergraduate course, ICCS will incorporate to this course lectures on the HPC benchmarking procedure, focusing both on tools and best practices. Further, ICCS will present the "Location graph extraction" component developed for the Migration Pilot.

Table 6. Training by ICCS on HPDA methods.

	Training by KNOW on the Analysis of Social Networks with AI
Curriculum Track	Both
Type / format	Online and on-site training
Aim	Provide Artificial Intelligence training to the consortium members and externals.
Venue & date	On-site or online, Y3.
Target group	Academia and industry representatives (with a strong technological/mathematical background), within and outside the consortium.
Training requirement addressed	Applying Artificial Intelligence [QUEST], Graph-analytics operations [HPDA_R1], Support for Python [TOOLS_R1].
Learned skills	KNOW currently offers an AI 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 leaders, innovation managers, process owners, and system administrators. In Y3, selected events will be adapted to the specific training needs of the consortium itself as well as external stakeholders, especially researchers, in terms of how methods from AI can be used to support the use cases in the HiDALGO project. Techniques for the "Analysis of Social Networks" will be demonstrated, to both create a general understanding of analysing networks, and to better understand the role of AI in supporting the SNA pilot. The training will make use of Python libraries for network analysis through Jupyter Notebooks in combination with the conda environment.

Table 7. Training by KNOW on the Analysis of Social Networks with AI.

	Training by PSNC on HPC, from introduction to optimisation
Curriculum Track	HPC and HPDA modelling
Type / format	Moodle material
Aim	Provide training to the consortium members and externals about basics of HPC systems, their usage and development.
Venue & date	Online and on-site, from Q3/2020.
Target group	Academia and industry representatives, within and outside the consortium.
Training requirement addressed	Introduction to HPC, Code optimisation for HPC, HPC using parallelisation [QUEST].

	Training by PSNC on HPC, from introduction to optimisation
Learned skills	As all pilots' codes are run in an HPC environment, it is necessary to provide basic knowledge about those systems to anyone intending to use them. PSNC online training utilises a computing infrastructure, where all applications, corresponding libraries, and tools necessary to demonstrate the pilots' functionalities are being deployed and configured. PSNC training will be offered on the HiDALGO online training system (Moodle) as well as live training, covering the following topics: ▸ 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. ▸ SLURM workload manager tutorial. ▸ Performing computations on HPC system. ▸ Resource management. ▸ Topological optimisation and parametric optimisation.

Table 8. Training by PSNC on HPC, from introduction to optimisation.

	Training by SZE on UAP modelling
Curriculum Track	Global Challenges
Type / format	Online and on-site training
Aim	Provide training to the consortium members and externals.
Venue & date	On-site (SZE) or online, Y3
Target group	Academia and industry representatives, within and outside the consortium.
Training requirement addressed	Applying Artificial Intelligence [QUEST].
Learned skills	In the course of Y2, SZE will prepare training for both developers within the consortium members, and for external stakeholders to be delivered in Y3. The UAP developers' training will focus on the details of the UAP workflows (data sources, databases, and interfaces). Aim of the training is to bring developers to connect functionalities of the HiDALGO infrastructure (related to visualisation, HPDA, and AI support) to deeply understand UAP, thus providing an opportunity to efficiently develop the CoE tools. The UAP users' training will aim at external stakeholders: city policy makers, environmental specialists, politicians, and civils who are interested in UAP modelling. The course will address the following topics: 1. Basic concepts of modelling, simulation, HPC, AI, and data analytics. 2. The HiDALGO infrastructure, tools, and the portal. 3. The air pollution challenge and the air quality directives of the European Commission. 4. The UAP workflows. 5. Hands-on tutorial on UAP simulations. Material for the Moodle platform will be prepared and made available as well.

Table 9. Training by SZE on UAP modelling.

	Training by USTUTT
Curriculum Track	Both
Type / format	Online and on-site training
Aim	Provide training to the consortium members and externals.
Venue & date	USTUTT or online, Y2 and Y3.
Target group	Academia and industry representatives, within and outside the consortium.

	Training by USTUTT
Training requirement addressed	Methods for the SNA pilot, HPC using parallelisation [QUEST], Support for Python [TOOLS_R1].
Learned skills (Global Challenges track)	HiDALGO plans to integrate original contributions in the following courses: ▸ Iterative Linear Solvers and Parallelisation: A lecture and a hands-on session on computing eigenvalue histograms (i.e., eigenvalue estimators) of large-scale sparse symmetric matrices with PETSc. This method is applied to large scale networks in the SNA pilot. ▸ CFD with OpenFOAM®: Current plans include a brief presentation of the UAP, in particular the results obtained with OpenFOAM®. The audience is then referred to the UAP HiDALGO spokespersons for further information and details. The March 2020 edition of this course had to be cancelled at short notice due to the COVID-19 outbreak.
Learned skills (HPC and HPDA modelling track)	▸ Parallel Programming Workshop (MPI, OpenMP and Advanced Topics). The MPI Python API is being added to the Fortran and C/C++ ones, to bring the participants closer to some packages extensively used within HiDALGO (MPI for Python [38], Dask.distributed [37]). Such an addition is a work in progress, and will be included in one of the parallelisation courses at USTUTT or at a GCS [17] location in Y2 or Y3. ▸ ML, DL, and GPU programming using OpenACC [35]: A DL course will be offered by NVIDIA in July 2020. A session by HiDALGO personnel on Data Compression of numerical data sets with the BigWhoop library and a hands-on session on the use of Apache Spark [33] through pySpark and Jupyter Notebooks are going to be included in this course. Further ML/DL courses are in preparation at USTUTT and scheduled for Y3.

Table 10. Training by USTUTT.

4.3 Summary of the technical topics in the two-track training curriculum

In Table 11 below, the two tracks of the training curriculum (HPC and HPDA modelling for GC scientists and analysts, and GC needs for HPC technology leaders) are mapped to the corresponding technical topics, covered by the planned and performed activities in Sections 4.1 and 4.2 above, as well as those included in the technical requirements in Section 2.1, but not yet implemented in specific training events. As already mentioned in Section 2.3, the two tracks complement each other, but at the same time some topics are common to the two tracks, which is beneficial for the exchange between the different communities.

Topic	Global Challenges Track	HPC-HPDA Track
Online project management and Agile Software Development (ASD) tools (e.g., OpenProject).	✔	✔
Tools for faster collective developers' work (e.g., Git and Jenkins).	✔	✔
Support for Python through Jupyter Notebooks.	✔	✔
Applying Artificial Intelligence, e.g., for the Analysis of Social Networks.	✔	✔
Graph-analytics operations (e.g., location graph extraction).		✔
Coupling of weather and climate (ECMWF) data.		✔
Coupling with both static and streaming external data sources (integration of streaming data into HPC environments).		✔
Coupling simulations across data-centres.		✔
Agent-based modelling and simulation tools (e.g., Flee).		✔
HPC benchmarking and optimisation procedures: tools (e.g., ScoreP and Vampir) and best practices.		✔
HPC: Introduction and workload management (e.g., SLURM), Code optimisation (topological and parametric optimisation), HPC using parallelisation.		✔
High-performance datastores and distributed file systems (e.g., Cloudify and CKAN).		✔
Support for high-performance, general-purpose analytics runtime.		✔
Parallel Programming with Python (MPI for Python).		✔
Apache Spark with pySpark and Jupyter Notebooks.		✔
Training on the UAP workflows (data sources, databases, and interfaces) for developers and users.	✔	
Training on the UAP with focus on CFD and OpenFOAM®.	✔	
Methods for the SNA pilot (e.g., eigenvalue estimators).	✔	

Table 11. Technical topics in the HiDALGO two-track training curriculum.

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. HiPEAC 2020 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 pre-processing phase, SZE has been constructing a complete educational course to be offered in Y2 (see Section 4.2).

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 most HPC systems). Furthermore, a dedicated synopsis 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 (Table 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 has been set up [40]. 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 [39] 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. HiDALGO training events have so far also been promoted through HiDALGO-internal and cross-CoEs mailing lists. Further channels are envisaged, such as an internal project Wiki and the Euro-HPC training registry for dissemination: Developments in this regard will be detailed in D7.4 [28].

Even though in-class events possess an incomparable effectiveness in transfer of know-how, web-seminars 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 web-seminars 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 [32]). 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 web-seminars, including presentations, hands-on sessions and discussions, is a promising start in this direction. Future digital contributions, such as the presentations and trainings provided at the multi-day workshops (Section 3.2), could serve as the basis for a Massive Open Online Course (MOOC).

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 (Table 1), where the training has been provided on site, with no fees for the participants. Also, compared to in-person training, web-seminars are an inclusive training practice, which drastically reduces participation fees and travel costs. Thus,

[1] Assessment of trainings and workshops of EXCELLERAT [12] Partners, February 2019, to be repeated in Q1 2020. See EXCELLERAT D5.1 [30] and D5.3 [31].

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 [30]). In Table 12, the targets for Y2 and Y3 are reported, together with the corresponding achievements (when available) 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%	75%

Table 12. KPIs for T7.3.

KPIs 7.3.1 and 7.3.2 have already far exceeded the targets set for Y2, and KPI 7.3.4 has also reached the target. In particular, all the participants (60) at the First Flee Workshop (Table 1) and at the HiPEAC event (Section 3.1) do not belong to the consortium, while those (19) who attended the two web-seminars (OpenProject and Git & Jenkins, Table 2 and Table 3) belong to either HiDALGO or EXCELLERAT [12].

On the other hand, KPI 7.3.3 could not be measured. This is due to the time needed to define the KPIs after the start of T7.3, and to realise the necessary tool for measuring quality. An online questionnaire has been finalised after the events considered in Table 12 and will be therefore distributed to the participants at the conclusion of each upcoming event. KPI 7.3.3 and the evolutions of the other KPIs will be reported in the next deliverable D7.4 [28].

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 12, 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 [30] and D5.3 [31]). So far, concerning the four events in Table 12, 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 [19] and Women in HPC [20]. 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 Coordination 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 the goal of "innovation workshops", training activities aim at fulfilling the HiDALGO two-track training curriculum for respectively HPC technology leaders and scientists and analysts who tackle GC through the support of HPC/HPDA. The specific HPC/HPDA needs arising from the pilots and the requirements of stakeholders and trainees must constantly be taken into consideration.

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 (Final Report on Community Building, Event Management, Collaboration and Training). 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 (Final Portal Release and System Operation Report).

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] 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].

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

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

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

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

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

[9] 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].

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

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

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

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

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

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

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

[17]Gauss Centre for Supercomputing, [Online] https://www.gauss-centre.eu/ [Accessed 2020-10-17].

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

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

[20]WHPC, [Online] https://womeninhpc.org/ [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]HiDALGO D4.2 – Implementation Report of the Pilot Applications. Horváth, Zoltán 2019.

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

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

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

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

[28]HiDALGO D7.4 – Annual Report on Community Building, Event Management, Collaboration and Training (due date 30/11/2020)

[29]HiDALGO D7.5 – Final Report on Community Building, Event Management, Collaboration and Training (due date 30/11/2021)

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

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

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

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

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

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

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

[37]Dask.distributed, [Online] https://distributed.dask.org/en/latest/ [Accessed 2020-10-17].

[38]MPI for Python, [Online] https://mpi4py.readthedocs.io/en/stable/ [Accessed 2020-10-17].

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

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

Annexes

A tabular overview of performed and planned activities building up the HiDALGO two-track training curriculum follows. For educational tasks, the shortcut of the corresponding track is indicated:

- Global Challenges [GC] needs for HPC technology leaders;
- HPC and HPDA modelling [HPC-HPDA] for GC scientists and analysts;
- Both [B]: Event of interest for both communities, or training on tools useful to the project or to other CoEs.

Date	Name of the event	Location	Type	Section
05/2019	International Symposium: Urban Systems – Global Challenges – Digital Tools	Stuttgart, Germany	Innovation Workshop	3.2
07/2019	First Flee Workshop	Adama, Ethiopia	Educational Task [HPC-HPDA]	4.1
08/2019	OpenProject Webinar (USTUTT)	Web	Web-seminar [B]	4.1
10/2019	Best Practice Guide for Git & Jenkins (USTUTT)	Web	Web-seminar [B]	4.1
01/2020	HiPEAC Conference	Bologna, Italy	Innovation Workshop	3.1
Q2/2020	USTUTT Web-seminar on CKAN	Web	Web-seminar [HPC-HPDA]	4.2
s05/2020	Forum TERATEC 2020	Palaiseau, France	Fair	3.2
06/2020	ISC High Performance 2020	Frankfurt am Main, Germany	Innovation Workshop	3.2
06/2020 (or 2021)	Multiscale Modelling and Simulation Workshop	Amsterdam, The Netherlands	Innovation Workshop	3.2
07/2020, Q3/2020, 2021	Deep Learning and GPU programming using OpenACC	Stuttgart, Germany	Educational Task [HPC-HPDA]	4.2

09/2020 (or 2021)	International Conference on Principles and Practice of Multi-Agent Systems - PRIMA 2020	Nagoya, Japan	Innovation Workshop	3.2
09/2020	Global Challenges: Risk and Complexity	Stuttgart, Germany	Innovation Workshop	3.2
09/2020	Social Simulation Conference	Milan, Italy	Innovation Workshop	3.2
09/2020 to 01/2022	ICCS HPDA Training in courses at the National Technical University of Athens	Athens, Greece	Educational Task [HPC-HPDA]	4.2
2020/2021	ECMWF HiDALGO-specific Training	tbe	Educational Task [B]	4.2
2020/2021	PSNC Training with PSNC Virtual Machines	Poznań, Poland	Educational Task [HPC-HPDA]	4.2
2020/2021	SZE Training on UAP	tbe	Educational Task [GC]	4.2
2020/2021	USTUTT HiDALGO-specific Training	Stuttgart, Germany	Educational Task [B]	4.2
2020/2021	Further Flee Workshops	Adama, Ethiopia / Web	Educational Task [HPC-HPDA]	4.2
Q1/2021	CoEs joint Winter School	Stuttgart, Germany / tbe	Innovation Workshop	3.2
06/2021	ISC High Performance 2021	Frankfurt am Main, Germany	Innovation Workshop	3.2
2021	KNOW HiDALGO-specific AI Training	tbe	Educational Task [B]	4.2
2021	HiDALGO-initiated Workshop	tbe	Innovation Workshop	3.2

Table 13. HiDALGO two-track Training Curriculum.