START PAGE

MARIE SKŁODOWSKA-CURIE ACTIONS

Innovative Training Networks (ITN)
Call: H2020-MSCA-ITN-2017

Synergies between Multivariate Analysis, Real Time analysis and Hybrid architectures for Event Processing
SMARTHEP

This proposal is to be evaluated as:
ETN

LIST OF PARTICIPATING ORGANIZATIONS

List of Beneficiaries

Consortium Member (beneficiaries)	Legal Entity Short Name	Academic	Non-Academic	Awards Doctoral Degrees	Country	Dept. Division Laboratory	Scientist-in-Charge
1. Lund University	ULUND	✓		✓	SE	Particle Physics	Dr. Caterina Doglioni
2. CNRS	CNRS	✓			FR	LPNHE	Dr. Vladimir Gligorov
3. TU Dortmund University	TUDO	✓		✓	DE	Faculty of Physics, E5	Dr. Johannes Albrecht
4. University of Heidelberg	UHEI	✓		✓	DE	UHEI	Prof. Dr. Stefanie Hansmann-Menzemer
5. University of Helsinki	UH	✓		✓	FI	Department of Physics	Dr. Mikko Voutilainen
6. EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH	CERN	✓			CH	Physics Department	Dr. Monica Pepe Altarelli
7. National Institute for Subatomic Physics, Nikhef	NIKHEF	✓			NL	Scientific Department	Prof. Olga Igonkina
8. Université de Genève	UNIGE	✓		✓	CH	DPNC	Prof. Anna Sfyrla
9. Sorbonne Université	SU	✓		✓	FR	LIP6	Dr. Lionel Lacassagne
10. IBM France	IBM		✓		FR	–	Christian de Sainte Marie
11. KKT	KKT		✓		IT	–	Lorenzo Taccari

Data for non-academic beneficiaries

Name	Location premises	Type of R&D activities	Full-time employees	Employees in R&D	Website	Annual turnover	Enterprise status	SME status
KKT	Firenze, IT	fleet intelligence	N	M	https://www.linkedin.com/company/kkt-srl/	XXX k€	YES	YES
IBM	Paris, FRA	algorithm and software development	7500	600	https://www.ibm.com/fr-fr/	2,15 B€	YES	NO

FTE = Full-time equivalent

List of Partner Organisations

Consortium Member (Partner)	Legal Entity Short Name (also used in this Document)	Academic	Non-Academic	Awards Doctoral Degrees	Country	Dept. Division Laboratory	Scientist-in-Charge	Role of Partner Organisation
Ximantis	XIMANTIS		✓		SE	–	Dr. Alexandros Sopasakis	Training, secondments on RTA in traffic prediction and decision-making
POINT8	POINT8		✓		DE	–	Dr. Kevin Dungs	Training, secondments on RTA in transport and logistics
Lightbox	LIGHTBOX		✓		CH	–	Dr. Pierluigi Catastini	training, secondment on RTA in IoT in industrial processes and financial decisions
University of Pisa and INFN	INFN	✓		✓	IT	Department of Physics	Dr. Alberto Annovi	Training, secondments in hybrid computing architectures for ATLAS
University of Santiago de Compostela	USC	✓		✓	ES	Department of Physics	Dr. Diego M. Santos	Training, secondments in GPU for RTA and physics for LHCb
University of Oregon	UO	✓		✓	US	Physics Department	Prof. David Strom	Training, secondment at CERN in hybrid computing architectures and physics for ATLAS
The Ohio State University	OSU	✓		✓	US	CCAPP and Physics Department	Prof. Antonio Boveia	training, secondment at CERN in FPGA, hybrid computing architectures and physics analysis for ATLAS
University of Liege	ULIEGE	✓		✓	BE	Department of Physics	Prof. Gilles Louppe	Training, secondment in machine learning for anomaly detection
VU University Amsterdam	VU	✓		✓	NL	Department of Physics	Prof. Gerhard Raven	Training, award of PhD degree
Radboud University Nijmegen	RU	✓		✓	NL	Department of Physics	Mario C. van der Toorn	Training, award of PhD degree
University of Bologna	UNIBO	✓		✓	IT	Department of Computer Science	Dr. Samuele Salti	Training, award of PhD degree

We declare no inter-relationship between different participating institutions or individuals.

B1. EXCELLENCE

The volume of data available to research and industry is ever-increasing. This acceleration in data collection is not always matched by a comparable acceleration in data storage and data utilization. This poses a **problem for the traditional paradigm of first collecting data, then analyzing it**. If the space to store the data is limited, then data has to be discarded or not recorded at all. Moreover, most of the data produced is recorded and stored without ever being analyzed, and loses value if insight is not extracted in a short period of time[1].

At the experiments at the Large Hadron Collider at CERN, information on fundamental components of matter from proton-proton collisions is provided up to 30 million times per second. A decision on whether a collision "event" is interesting must be made in the order of milliseconds, and novel techniques are needed in order to extract information that would otherwise be discarded. For commercial applications such as mobile apps for fleet control and transport optimization, IoT sensors and fraud detection, the amount and complexity of the data grows while the resources and time to take a decision using that information do not scale accordingly – industry faces a similar challenge as High Energy Physics (HEP).

In order to make the most of the available information in a cost-effective way, and to reduce the time between data collection and insight into the results to inform subsequent decisions, data-taking and data-analysis need to be fast and efficient. This is why SMARTHEP (**S**ynergies between **M**ultivariate **A**nalysis, **R**eal **T**ime analysis and **H**ybrid architectures for **E**vent **P**rocessing) is an interdisciplinary ETN aiming to train Early Stage Researchers (ESRs) on **Real Time Analysis (RTA)** techniques, where **data collection and analysis are nearly simultaneous**, using novel software and hardware tools. Within SMARTHEP, the ESRs will employ machine learning (ML) and multivariate techniques as well as artificial intelligence (AI) to speed up the reconstruction and the analysis of the data, introducing novel algorithms that improve reproducibility and accountability of real-time decision making. The ESRs will become experts in novel hybrid software/hardware solutions, where data can be collected and reconstructed on the same platform.

The SMARTHEP **ESRs will be trained** by a consortium of successful researchers in **High Energy Physics and Computer Science** (including 5 holders of ERC StG and CoG, as well as recipients of multiple national grants, and includes a mentoring group of experienced scientists worldwide) and entrepreneurs from industries ranging from multinational to start-up. These scientists, professionals and entrepreneurs have begun their collaboration within SMARTHEP and will continue working together beyond this action with a commitment to follow the future careers of the scientists they train. The ESRs will **learn transferrable skills** from the exposure to research, industry and society and network events. They will be **prepared for careers in industry or science** and **contribute to European growth** from their hands-on work on concrete commercial and research deliverables, such as mobile apps using ML for efficient vehicle fleet control, optimized algorithms for traffic and public transport predictions, software for sensors that will monitor industrial production, accountable and understandable algorithms for fraud detection and decision-making in finance, as well as cutting edge computing technologies and algorithms necessary for the detection and measurement of fundamental particles and for the discovery of new physics phenomena.

B1.1 Quality, innovative aspects and credibility of the research programme

B1.1.1 Introduction, objectives and overview of the research program

Researchers in HEP study the basic constituents of matter and their interactions within the theory called the Standard Model (SM), and search for new physics phenomena explaining the SM's shortcomings. One of the main experimental tools available for this study is the LHC. At the LHC, particle beams are accelerated to high energies and collided every 25 ns. The products of the collisions are analyzed by scientists operating four main particle detectors: ALICE, ATLAS, CMS and LHCb[2]. Collision *events* occur 30 million times per second. Recording each event would be prohibitive for the current data acquisition and storage systems. Since many of those events contain known processes that are not considered interesting, each experiment has a *trigger system* composed of hardware and software. The trigger system analyzes collision events in real time, with a maximum delay of micro- to milliseconds, and selects which ones to keep for further analysis. This real time event selection is crucial for HEP, especially after the LHC second running period (Run-2) has drawn to a close. While there are many observed phenomena that are still not included in the SM (e.g. dark matter [LINK BOVEIA DOGLIONI ANN REV]), the leading predictive theories such as Supersymmetry have so far yielded null results. This is prompting HEP researchers to reconsider their definition of "interesting". New phenomena could have unexpected, not-yet-theorized manifestations in the detector, and they would be missed using the current trigger algorithms. New phenomena could also be extremely rare and buried in uninteresting backgrounds, and require specialized data taking techniques. Moreover, disagreements of data with the SM prediction require dedicated investigation with the new dataset, and therefore dedicated

[1] W. Riess, IBM Zurich, Nanoscience Colloquium at Lund University titled "The Future of Computing", 17/05/2018 [2] CERN, The LHC Experiments, ~~CERN Document Server~~

trigger reconstruction and selection criteria. For example, the SM prediction that the weak force couplings to all lepton types are equal (Lepton Flavor Universality, or LFU) has recently been challenged by LHCb measurements [3]. This makes LFU and LFV (Lepton Flavor Violation, the prediction that the overall number of leptons of a given type does not change in interactions) among the most interesting topics for the near-future of particle physics and one that needs RTA to be explored.

As detailed in Secs. 1.3.1 and **??**, researchers in the SMARTHEP consortium are experts on the trigger systems of all four main LHC experiments. Within SMARTHEP, they will pool their expertise and complement it with that of computer scientists, entrepreneurs and professionals.

SMARTHEP will train a new generation of researchers with a strong expertise on RTA techniques and analysis of large datasets. The ESRs will exploit innovative ML techniques and hybrid computing architectures to upgrade the trigger systems of LHC experiments for the next LHC data taking periods where the volume of data will increase even further. They will perform physics analysis with these improved systems, towards precise probes of the SM and searches for new phenomena.

SMARTHEP's ESR projects employ RTA in research as well as in projects that are useful to industry and society. Time-to-insight from raw data is a crucial benchmark for a competitive and effective use of the information. For this reason, advancement in terms of fast and efficient data analysis and decision-making are required by society and the commercial sector, and a selection of industrial use cases sharing common issues with HEP has been chosen as integral part of the SMARTHEP research and exploitation program as described further in Sec. **??**. RTA is a key point for improving transport and its safety, as well as for optimizing industrial processes. The information on the position of private or public transport vehicles and their possible routes are transmitted to a central analysis system, so that the traffic conditions around it are analyzed and a forecast is made. All of this needs to be done on a timescale that is short enough for a city to modify its traffic light system to avoid congestion, or to reroute other vehicles to avoid delays. This requires both ML and hybrid computing architectures for algorithm training, similarly to the solution of complex problems such as tracking particles in a HEP detector. The detection of a driver's behavior (e.g. drowsiness) in fleets of cars needs to happen in the timescale of milliseconds to preserve safety, and detection of malfunctioning components in industrial processes must trigger maintenance before the component breaks. Both those use cases require the optimization of advanced algorithms on limited computational resources, much like HEP's trigger systems. Another use case for RTA in industry is fraud detection, where the detection of an anomalous transaction and a decision on whether it is allowed or not needs to be made on a short timescale. This is a similar challenge to HEP triggers looking for unexpected phenomena. While ML algorithms trained to detect outliers are the current state of the art for this purpose, their black box approach raises interpretability and explainability challenges. This is one of the topics beyond the state-of-the-art that is tackled by SMARTHEP: knowledge-based approaches for the induction of explicit decision rules are useful both for increasing the accountability of fraud detection algorithms and for HEP trigger systems where every decision to discard an event is final and must be completely understood.

The cross-talk on RTA between academia and industry in SMARTHEP will lead to concrete deliverables that contribute to improving European transport as per the H2020 Societal Challenge for Smart, green and integrated transport, to enhancing the capabilities and reproducibility of decision-making in finance and to improving industrial processes.

Consortium composition

SMARTHEP spans ten different countries. CERN, the NIKHEF and CNRS research organizations, and universities SU, UH, UNIGE, TUDO, ULUND, UHEI form the core academic side of the Network, with experts in HEP and computer science. IBM and KKT as beneficiaries, and XIMANTIS, LIGHTBOX, POINT8, as partners, form a balanced counterpart to academia with a focus on the same technological challenges as HEP to gain insight into data in real-time. UO, OSU, INFN, USC, ULIEGE, RU and VU are high-profile associated academic partners whose expertise reinforces the network, that provide training and secondment to the ESRs, and award PhD degrees in case the beneficiaries are industries and research institutes [4]. All consortium members have successfully proven their capabilities for research, training, exploitation and dissemination as shown in Sec. **??**, **??** and part B2.

[3] LHCb Collaboration, Test of lepton universality with $B^+ \to K^+\ell^+\ell$, Phys. Rev. Lett. 113, 151601 (2014). LHCb Collaboration, Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell$, JHEP 08 (2017) 055. [4] In this proposal, we use "node" to indicate either beneficiary or partner. The *node responsible* is the *main scientific contact person* for the node. Administrators, albeit indispensable for the network functioning, are not named even if they are node responsibles in the case scientists are affiliated to two nodes.

Network goals and research objectives

SMARTHEP trains researchers in ML, AI and data analysis and hybrid computing architectures, in order to enable RTA for the advancement of decision-making as well as monitoring and discoveries in both research and industry. The work and expertise developed within SMARTHEP will develop and commission upgraded triggers systems for LHC experiments for HEP measurements and searches, improve transport, industrial process and financial decision-making. These network goals define the four main research topics for SMARTHEP, each corresponding to a research Work Package (WP) detailed in Sec. 1.1.2. Each of the research topics has research objectives (RO) shown in each table. The RO can be achieved with this network thanks to the collaboration between industry and HEP, and they are linked to concrete deliverables.

The first research topic concerns the use of **ML, AI and advanced data analysis** in RTA techniques. ML techniques are by now ubiquitous in both HEP and industry, in order to optimally analyze raw, unstructured data. In SMARTHEP we will further extend their use in RTA and event reconstruction, paying particular attention to their understandability and reproducibility (see research topic 4). Meta-analysis (benchmarking) is also necessary to understand the algorithm performance prior to using them in RTA and to satisfy he requirements of resource-constrained environments.

Research Objective:	**Outcomes:**
design and use a variety of ML algorithms for RTA (GAN, RNN, Deep Learning)	HEP and commercial software toolkits (ADD DELIVERABLES);
implement efficient real-time object and event reconstruction	software for HEP experiment triggers (ADD DELIVERABLES);
benchmark and optimize RTA algorithm performance	inter-sector benchmarking toolkits (ADD DELIVERABLES);

The second research topic focuses on the design of innovative **hybrid computing architecture** solutions in hardware/software, given that the complexity and rate of LHC data and beyond does not allow standard processors or data analysis techniques to be competitive[5]. An example of such a technology in HEP is the Fast TracKer in ATLAS, a hardware-based ATLAS Fast TracKer (FTK)[6], a unique hardware processor made by several custom electronic boards based on Field Programmable Gate Arrays (FPGA) to reconstruct the trajectory (track) of the charged particles that cross the inner part of the experiment. The FTK is a complex system made by several custom electronics boards based on FPGAs and unique computing devices. In industry, Graphic Processing Units (GPU) are used for computing parallel processes and training ML algorithms, and their use is being prototyped in HEP by members of SMARTHEP.

Research Objective:	**Outcomes:**
use of FPGAs for fast hardware computations	trigger improvements for LHC experiments (ADD DELIVERABLES);
use of GPUs for massively parallel computations	significant speed-up for HEP and industrial ML and RTA techniques (ADD DELIVERABLES);
design parallelized and multithreaded algorithms	RTA software needed for the increase in LHC data and for more efficient financial transactions (ADD DELIVERABLES);

The third research topic concerns **real-time decision making in physics and industry**. The outcomes of the previous research topics will be applied to concrete physics cases and industrial challenges. These improvements in RTA benefit each experiment's physics program and increase the information quality extracted from data-rich environments in industry. There are many HEP drivers behind the trigger improvements that will be the SMARTHEP ESR's thesis topics: hints of physics beyond the SM in LFV/LFU, the mystery of 85% of the matter content of the universe that is not yet included in the SM, the in-depth study of the newly discovered Higgs boson and the study of collective particle behavior from states of matter created shortly after the Big Bang. Similar techniques used for trigger improvements will also help improve transport and fleet safety.

Research Objective:	**Outcomes:**
deploy improvements from real-time reconstruction and hybrid architectures in HEP trigger systems	measurements and searches with LHC Run-3 data (ADD DELIVERABLES) ;
use ML and hybrid architecture for RTA to improve transport and safety of fleets	inputs to traffic management of connected cities, extended transport apps and in-vehicle software (ADD DELIVERABLES);

The fourth research topic concerns efficient and accountable **monitoring** of complex processes and increase the sensitivity to physics **discoveries** through outlier detection and novel RTA techniques in data taking. Firstly, industry and HEP share the challenge of not knowing in advance what kind of "interesting" events will appear, in

[5] J.P. Vlimant, Machine Learning in Charged Particle Tracking, Hammers and Nails ML&HEP Conference, 2017 [6] J Adelman et al., ATLAS FTK Challenge: Simulation of a Billion-fold Hardware Parallelism, Procedia Comput.Sci. 66 (2015) 540-545

the case of malfunctioning equipment, fraudulent financial transaction or potential not-yet-theorised new physics process. Outlier/novelty detection is a growing field, with dedicated ML algorithms[7]. In SMARTHEP, experts in the field from IBM and their supervised ESRs will apply those to physics and industry use cases, and extending them to methods where rules on what are inferred from a mixture of theory and known events to increase their accountability and understandability. Secondly, the present HEP paradigm, in which a pre-selection of interesting signals based on simple event features reduces the event rate to a manageable size for further storage, works well in an environment dominated by uninteresting backgrounds. However, crucial HEP research questions cannot be answered entirely by physics processes with these characteristics. Not being able to fully exploit the wealth of collision data delivered by the LHC because it is impossible to save it for further analysis is a severe cost for the advancement of the physics program of all LHC experiments, especially in light of the planned upgrades to the LHC collider that will significantly increase the collision rates. For this reason, many of the members of SMARTHEP have been extending the physics program of their experiments using novel RTA techniques in data taking[8]. The technique they pioneered exploits real-time reconstruction and analysis at the trigger level to only record high-level information, rather than the raw detector information as it is traditionally done. Since high-level quantities are much smaller than raw data, more events can be recorded and used for physics analysis, and and a much increased sensitivity for measurements and discoveries can be achieved[9]. SMARTHEP will improve and extend this technique beyond the current state-of-the-art towards the upgrades, in LHCb where most of the data will be recorded this way, in ALICE where it will be implemented for the first time, and in ATLAS and CMS where it will be applied to a larger number of searches.

Research Objective:	Outcomes:
implement outlier detection techniques in physics and industry	potential for not-yet-theorized discoveries at the LHC, improved fraud detection algorithms and IoT-ready industrial predictive maintenance (ADD DELIVERABLES) ;
study and design rule-induction algorithms as complement to ML to improve accountability and understandability	RTA for fraud detection and HEP triggers (ADD DELIVERABLES);
extend RTA data-taking techniques in HEP trigger systems where raw data is discarded	increased sensitivity to new physics processes in LFV/LFU, Higgs boson measurements and dark matter searches

B1.1.2 Research methodology and approach

SMARTHEP defines three WPs for the management of the consortium, for training, and for outreach and dissemination, and four research WPs, corresponding to each of the research topics in the previous section.

In **WP3**, we make full use of advanced **Machine Learning, AI and data analysis** techniques in RTA. This is an emerging field of research that enables real time decision-making, monitoring and discoveries in industry and society. All ESRs will be involved in this WP, including studies of best practices and performance (ESR11, ESR12).

In **WP4**, we will develop algorithm implementations which efficiently use modern computing hardware that is increasingly both highly parallel (ESR3) and heterogeneous, with CPU-based processing farms complemented by GPU or FPGA co-processors (ESR6, ESR7, ESR13). This kind of **hybrid architectures** enables faster data processing, and their design requires highly-specialized training which few contemporary researchers in the physical sciences and that the consortium including computer scientists is best placed to deliver.

In **WP5**, we will employ the techniques and tools developed in WP1 and WP2 to enable advanced RTA techniques for **decision-making**. In HEP, we will develop and commission the upgraded triggers of the main LHC collaborations and use them for physics analysis (all ESRs), while in industry we will improve traffic and fleet safety (ESR4, ESR11, ESR13, ESR15).

Within **WP6**, we will focus on **real time monitoring of complex processes and preparations for unexpected discoveries**. This feeds back to the decision-making WP5 through novelty/anomaly detection algorithms (ESR9) that are understandable and reproducible, in HEP, finance (ESR8) and predictive maintenance (ESR2). We will perform physics analysis that is fully based on RTA without the need to save raw data (ESR1, ESR4, ESR14, ESR13, ESR10).

All results will be **disseminated and communicated** to a variety of audiences within **WP7**, paying particular attention so that publications, software and documentation are accessible as legacy of the action according to the FAIR principles.

[7] R. Domingues et al, http://doi.org/10.1016/j.patcog.2017.09.037, Pattern Recognition 74 (2018) [8] e.g. Gligorov and Fitzpatrick, Anatomy of an upgrade event in the upgrade era, and implications for the LHCb trigger, CERN Document Server. ALICE Collaboration including Christiansen, Technical Design Report, CDS Document Server, ADD ATLAS TRIGGER PAPER [9] CMS Collaboration including Pierini, Search for narrow resonances in dijet final states at $\sqrt{s} = 8$ TeV with the novel CMS technique of data scouting, Phys. Rev. Lett. 117, 031802 (2016). ATLAS Collaboration including Boveia, Starovoitov, Dunford and Strom, ADD HERE TLA REF, LHCb collaboration including Gligorov, Borsato ADD HERE DARK PHOTON REF

HEP experiments are by their very nature built around **training**, as around a third of all LHC collaborations consist of PhD students. The training WP (WP2) encompasses WP3-7 and combines this collaborative, training-based research culture, with a focus on cutting-edge technologies.

Throughout this ETN, RTA will provide the common language linking these challenges and disciplines, in research, academia and industry, and will enable each of the WPs to benefit from the progress made in the others. The WPs are summarized in the Table below, and Fig 1. gives a schematic view of the four research WPs and the main actors within the consortium.

WP No.	WP Title	Lead Beneficiary No. / responsible	Start	End	Activity Type	Lead Beneficiary Short Name	ESRs Involvem't
WP1	Management	N, Doglioni	1	48	Management	ULUND	-
WP2	Training	N, Sfyrla	5	42	Recruitment, training	UNIGE	All
WP3	ML, AI and data analysis	N, Gligorov	1	48	Research	CNRS	1-14
WP4	Hybrid architectures	N, Lacassagne	1	48	Research	SU	2,3,6,7,13,14,15
WP5	Real-time decision making	N, Albrecht/Sopasakis	1	48	Research	TUDO	All
WP6	Monitoring & discoveries	N, de Sainte Marie/Pierini	1	48	Research	IBM	1,2,4,6,8-10,13,14
WP7	Dissemination & communication of results	N, Petersen	1	48	Dissemination, outreach	CERN	All

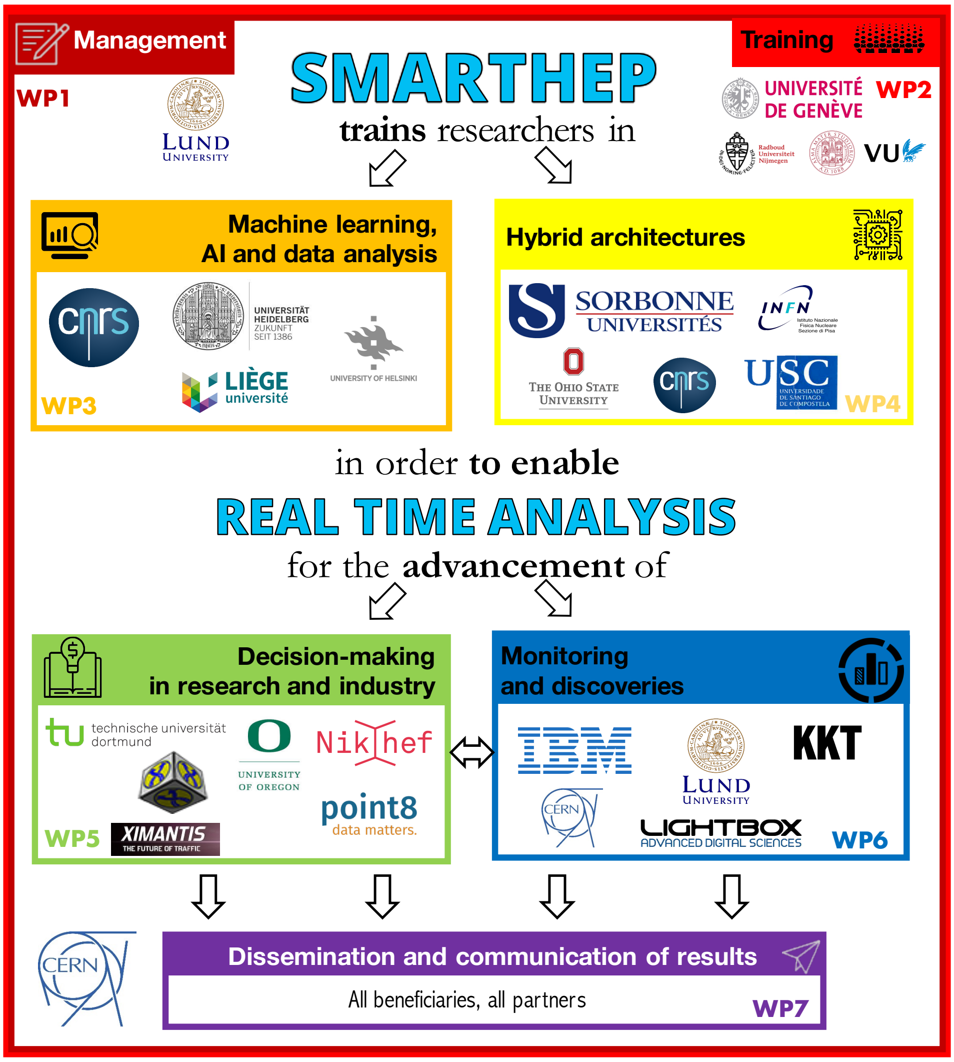

Figure 1: SMARTHEP implementation strategy and main node expertise.

In SMARTHEP, cutting-edge techniques in ML and hybrid architectures will be assessed and developed by the ESRs. They will be applied to specific use cases to RTA techniques that advance HEP, industry and society, and the results exploited, disseminated and communicated (Sec. ??). These steps correspond to the the milestones in Sec. . This modus operandi is at the heart of SMARTHEP and ensures the efficiency of its implementation.

B1.1.3 Originality and innovative aspects of the research program

The ESRs in the SMARTHEP network will undertake a timely and ambitious research program on RTA, benefitting from the experience not only of their direct supervisors, but of the entire network, to reach their objectives. Some of the original and innovative aspects of the projects in the network are listed below.

1. Researchers from SMARTHEP process large datasets using RTA techniques in a cost-effective way. A part of the novelty of this proposal is the volume of data which will be processed in real-time, comparable to the largest commercial tasks, but on a much smaller budget. One can compare Facebook and e.g. the LHCb collaboration. The former spends hundreds of millions of dollars a year on computing[10] while the latter only spends order 10M dollars a year on computing [11]. To achieve this goal and maintain a cost-effective data taking strategy, LHC experiments need a more systematic application of RTA (especially fully RTA-based analysis that discards bulky raw data), machine learning and hybrid architectures for HEP. As RTA is essential in industry as well, the research environment can benefit from a generation of ESRs with experience in industrial grade algorithms and tools. In particular, the ESRs will be trained in topics matching the emerging trend of real-time big data analytics[12] by invited speakers from large enterprises using these methods.

2. The SMARTHEP research program could lead to breakthroughs in our understanding of nature Only fully-real time data acquisition techniques and trigger improvements enable full exploitation of the LHC dataset. The HEP research topics chosen to drive conceptual developments within SMARTHEP have the potential to lead to the discovery of new physics beyond the Standard Model. Examples include studies of lepton flavor and its conservation and universality in different final states and experiments, dark matter mediators and new light particles, precise probes of the Higgs boson and of heavy ion physics, as well as not-yet-theorized phenomena that would not be possible to record using the current state-of-the-art triggers.

3. ESRs in SMARTHEP deploy and disseminate their research at a unique time for particle physics. As highlighted by the HEP Software Foundation Whitepaper[4], the period 2019-2023 is ideal for this R&D in HEP, as it is a time of transition between LHC data taking periods that will be necessary to prepare for an upgrade of the LHC accelerator where the amount of data delivered will make RTA techniques the key to pursue the physics programs of the four main LHC experiments. The systematic optimization of HEP experiments by SMARTHEP will boost the performance of the current and planned upgrades of the CERN based accelerator experiments. Furthermore, the developed toolkits will be advertised at international conferences and thus the developed methods will shape the online event selection of all future HEP experiments.

4. SMARTHEP researchers further the RTA paradigm in both HEP and industry. The close links of the research institutes of the consortium with the industry partners means that the ESRs will directly drive the development of novel industrial products, while also bringing professional methods of data mining that are exercised in large companies into the academic environment. Most modern methods applied in research can in this way be transferred to industry applications. By exposing industry-grade methods to the volume and complexity of HEP data, we will stimulate their development in a complementary way for the benefit of industry.

YR SUGGESTED PAGE COUNT - KEEP TO 7-8+4 PAGES UP TO HERE

B1.2 Quality and innovative aspects of the training programme

The training program of SMARTHEP has three main **objectives**, taking the Salzburg Principles[13] as a guidance:

- **Objective A**: provide ESRs with knowledge and training to conduct original research during and beyond their PhD studies;
- **Objective B**: provide solid bases in a broad spectrum of topics related to their field of research, extending beyond their dedicated research and including soft skills;
- **Objective C**: provide up-to-date and career-related training through interactions with multiple collaborators within academia as well as the industry, to meet the needs of a broad employment market.

Objectives A and B will be fulfilled by schools, workshops and events organized within the network; ESRs will profit from high quality lectures delivered by experts within the institutes associated to SMARTHEP, complemented by doctoral training at local nodes. Objective C will be fulfilled by special network events where the industrial collaborators will provide seminars and training on commercial applications of SMARTHEP's objectives.

[10] Facebook, The Facebook Data Center FAQ, 2010. [11] Private communication, Prof. Peter Clarke, University of Edinburgh. [12] A comprehensive guide to real-time big data analytics. [13] Salzburg II recommendations, http://www.eua.be/Libraries/publications-homepage-list/Salzburg_II_Recommendations

All these objectives will be further addressed via the natural collaboration between the ESRs and the various institutes of the network, as well as the planned secondments in academia and industry. This is a more flexible and diverse structure compared to standard PhD studies and will allow the ESRs to develop an individual and unique research mindset within an inclusive environment, so that they can act as proactive participants in furthering HEP and industry goals through RTA techniques.

B1.2.1 Overview and content structure of the training

Training is at the heart of all activities within SMARTHEP, with a dedicated Work Package, WP2. ESRs will have clear recruitment deliverables, a Personal Career Deevelopment Plan (PCDP), and they will be able to attend complementary network-wide and local training accounted through a credit system designed for this ETN.

Table 1.2.1: Recruitment deliverables per beneficiary, and supervisors/tutors with number of supervised Master's, PhD students and post-docs in brackets. The main supervisor, secondment supervisors and industrial tutors form the Supervision Committee (SC) of each ESR. Non-academic supervisors and tutors are in italics.

ESR	Recruiting Participant	PhD-awarding Participant	Planned Start	Duration	Main supervisors	Secondment supervisors and industrial tutors
ESR1	UH	UH	8	36	Voutilainen [9]	Pierini [17], *Taccari*
ESR2	UNIGE	UNIGE	8	36	Sfyrla [4]	Martinez-Santos [N], *Catastini*
ESR3	CERN	UNIGE	8	36	Petersen [6], Sfyrla [4]	Crescioli [4], *Catastini*
ESR4	TUDO	TUDO	8	36	Albrecht [13]	Martinez Santos [N], Matev [N], *Sopasakis* [20]
ESR5	TUDO	TUDO	8	36	Albrecht [13]	Raven [N], *Dungs*
ESR6	CNRS	SU	8	36	Crescioli [4], Malaescu [4]	Roda [14], *Sambo*
ESR7	SU	SU	8	36	Lacassagne[13]	Petersen [6], Couturier [N], *Catastini*
ESR9	IBM	ULUND	8	36	*Jullié*, Doglioni [11]	Louppe [N], Boveia [4]
ESR8	IBM	SU	8	36	*Feillet*, Lacassagne [13]	Gligorov [5]
ESR10	ULUND	ULUND	8	36	Christiansen [17]	Shahoyan [7], *Sopasakis* [20]
ESR11	NIKHEF	RU	8	36	Igonkina [14]	Strom [N], *Sopasakis* [20]
ESR12	NIKHEF	VU	8	36	Raven [14]	Albrecht [13], Petersen [6], Couturier [N], *Brambach*
ESR13	UHEI	UHEI	8	36	Starovoitov [7], Dunford [8]	Strom [N], Annovi [N] *Sopasakis* [20]
ESR14	UHEI	UHEI	8	36	Hausmann-Menzemer [N]	Albrecht [13], Martinez-Santos [N], textitDungs
ESR15	KKT	UniBo	8	36	*Sambo* [N], Salti [N]	Lacassagne [13], Doglioni [4]
Total			540			

Recruit deliverables per participant and awarding of PhD degrees. Table 1.2. present the recruitment deliverables per participant. ESRs will be recruited by month 8, complete a total of 36 months of research and training, and be awarded a PhD degree. ESRs at non-academic beneficiaries and ESRs at international or national laboratories will be awarded PhDs by universities within the network. For Finland, Netherlands and Sweden which mandate a four-year PhD, all beneficiaries will provide support for the ESRs to complete their PhD thesis.

Personal Career Development Plan. Two months after the start of their projects, ESRs and their SC will present a **Personal Career Development Plan** (PCDP) so that the **core, advanced and transferrable skills** to be acquired, as well as the milestones for each of the ESR projects, can be agreed between student and supervisor and consortium, planned and monitored throughout the course of the program, taking into account the existing resources both at the ESR node and at foreseen secondments. The PCDP will be reviewed by the local node coordinator, who will bring it to the Supervisory Board (SB) for approval. The main supervisor will be available throughout the course of the PhD, and meet with the ESR on a weekly basis and with an open-door policy. The local node coordinator will also review the progress at minimum every six months (e.g. during staff appraisal meetings) and bring a short report to the consortium. The PCDP[14] will include the requirements, milestones and goals within the schedule of the doctoral program (including secondments), the local and network-wide courses and schools to be attended, and a list of dissemination, communication and outreach activities.

Network-wide events: schools, yearly meetings and schools.

[14] The SMARTHEP PCDP will follow a common template based on that from the MSCA website to be coherent between the ESRs.

Network-wide schools, conferences and events shown in the table below will be organized by SMARTHEP beneficiaries and partners as part of the training and dissemination program and its preparation. We expect ESRs to attend network events in person, but wherever possible will make network school, conferences and events available as Webinars using the Vidyo technology provided by CERN, to allow all SMARTHEP ESRs and PIs to attend if family/personal commitments would otherwise prevent it. A permanent record of the lectures will be available as proceedings, and in some cases video recordings, as SMARTHEP has the ambition to make the training program available beyond this Action and continue organizing successful schools. The table below summarizes all events included within the network, with compulsory schools marked in bold so that students attend a yearly meeting and a school each year, dedicating sufficient time to local training and research. The hosts and lecturers for these events schools have been identified within the network based on their expertise, see Sec. B1.4.1.

	Main training events and conferences	Credits	Lead (support) institution	Action month	Notes
1.	**Kick-off meeting**	-	ULUND	2	-
	The kick-off meeting will be dedicated to organizing the project management, signing the consortium agreement, and monitoring the ESR recruitment.				
2.	**Yearly conferences**	2	ULUND (NIKHEF, IBM)	9,23,36	-
	SMARTHEP will hold **yearly in-person network-internal conferences**, with a duration of 4 or 5 days. On the first two days, the ESRs will gain experience in presenting their research by giving presentations to the other SMARTHEP members and participating in a poster session. Days 3–5 will be dedicated to lectures on research, technical and transferable skills, as well as communication and dissemination activities tailored to local circumstances (see Sec. ??). The first yearly meeting on month 8 will serve as introductory school on HEP and RTA, as it will be hosted shortly after all ESRs have been recruited. All yearly meetings will include a dedicated half-day of lectures on non-academic training. Within the Yearly Meetings, there will be time devoted to management activities such as the Executive Board, ESR Board and Supervisory Board meetings (see Sec. ??), that also train ESRs in scientific collaboration and governance.				
5.	**Non-academic training workshop**	1.5	ULUND	10	Network event joint with INSIGHTS ETN (TBC), compulsory for ESRs
	This workshop will follow the first yearly meeting. In this workshop, the ESRs will receive non-academic training by local experts at ULUND and the University of Oslo. The topics of the lectures will be diversity and inclusion, team-work, research integrity and sustainable research, as an early complement to local training on transferrable and soft skills so that the ESRs can start their work in a positive environment. The joint organization of ULUND and the University of Oslo, both within the INSIGHTS ETN (where C. Doglioni is a diversity and inclusion officer), will allow ESRs from two different ITNs to meet, exchange ideas and experiences, and broaden their network.				
4.	**Physics and machine learning school**	3	UNIGE	15	New network event, compulsory for ESRs
	This school will provide all ESRs with more advanced courses on the physics topics tackled in the network, a few months after the ESRs have started their projects. It will also provide an introduction to how to design a physics analysis, as well as to machine learning concepts and their connections to HEP and industrial applications.				
5.	**International School of Trigger And Data AcQuisition (ISOTDAQ)**	3	UO	18 or 30	-
	ISOTDAQ is a yearly school dedicated to triggering and acquiring data for physics experiments with lectures and hands-on exercises in equal proportions. A lecture about RTA by UO members of the SMARTHEP network will be added to the program if funded. ESRs will be encouraged to follow one of the two editions of the school during their PhD.				
5.	**Machine Learning for HEP school (MLHEP)**	3	CERN	22 or 34	-
	MLHEP is a school on cutting-edge machine learning techniques featuring dedicated trigger lectures, whose main organizer (Ustyuzhanin) is attached to CERN as a LHCb member. [Add some text] ESRs will be suggested to follow one of the two editions of the school during their PhD, and the nodes in SMARTHEP will be encouraged to place a bid to host the school as ULUND did in 2016.				
6.	**Basic FPGA course, FPGA boot-camp**	3	OSU, INFN	26, 27	New network event
	This school will include lectures on technologies and architecture and hands-on exercises on triggering applications. This school will consist of introductory courses given at CERN, and a follow-up bootcamp in Pisa where practical problems are solved for ESRs specializing in this topic. Lectures will be given from researchers in OSU and INFN, both leading institutes in research and development on hardware track triggers.				
7.	**GPU and hybrid architectures school**	3	USC (SU)	29	New network event
	In this school, the ESRs will learn how to compare architectures and practical solutions, and programming on different platforms (e.g. GPU programming). The ESRs will learn not only about what is available on the market and what is being planned, but also ways to best evaluate the chosen hardware solution for the software they are developing.				
8.	**Industry and career development school**	3	TUDO (XIMANTIS)	38	New network event
	This school will include lectures and workshops in collaboration with industry, and it is optimally placed towards the end of the ESR's PhD training in month 44. This school is dedicated to an in depth study of strategies for intellectual property rights (IPR), commercializing research output, presenting research results to policy-makers, and knowledge transfer. One day of this school will be dedicated to group work on case studies prepared by the industrial beneficiaries and partners. This school will include experiences and Q&A sessions with the CEOs and founders of the companies within SMARTHEP.				
9.	**Final public conference**	2	CNRS	42	Yes
	We will hold a five-day conference which will showcase the work of the network to the wider scientific community. As opposed to the yearly meetings, the conference will not feature management meetings, and will be dedicated to presentations on SMARTHEP research. Invited topical presentations on state-of-the art developments within the HEP and industry. Each of the conferences will dedicate between half and one day to dissemination activities by SMARTHEP members. The conference will have two additional days reserved for ESRs' presentations and public lectures showcasing the work done during their PhD program with SMARTHEP. ESRs will also take active part in the organization of this conference, to add to their transferrable skillset.				
10.	**Closing meeting**	-	UHEI	48	
	In the closing meeting, the network will take stock of the experience of the ETN and plan the next steps. The closing meeting is beyond the doctoral period for most of the ESRs, but they will be invited to participate as network alumni, also to give feedback on the ETN experience. The PIs and the ESR alumni will give public lectures.				

All network-organized schools will be also open to the local students of the beneficiary and partners organizing the school and advertised through the HEP Software Foundation Training group, as part of enhancing the overall

training program of the institutions involved in SMARTHEP and broaden participation to the network's activities. External students will be accepted if the optimal capacity of the school is not reached by the participating ESRs alone, and a fee will be charged to non-network participants if the node incurs external expenses due to their participation.
Local training. Training events organized specifically for SMARTHEP are complemented by local training provided by beneficiary nodes. Students from the network will be able to participate in this training when located at the node. All nodes include a range of graduate level courses in languages, career management, presentational skills, diversity and inclusion, as well as pedagogical courses in teaching and learning that the ESRs will be encouraged to follow to obtain the necessary amount of credits. Partners will also contribute with individual supervision and local training for employees while students are seconded at their premises. Links to the training courses for each node are detailed in their description, while below we list examples of graduate schools or training programs within the academic nodes that the ESRs will be associated to.

Beneficiary	Graduate school or program
ULUND	COMPUTE graduate school in advanced computing techniques for research
CNRS	Add something here
TUDO	Add something here
UHEI	HighRR graduate school on HEP detector development.
UH	HEP graduate school (40 ECTS points, of which 10 of transferrable skills)
CERN	Academic training program including transferrable skills.
NIKHEF	Research school in sub-atomic physics.
UNIGE	HEP doctoral school in theory and phenomenology.
SU	Add something here

As SMARTHEP also intends to prepare ESRs to teach others, specific credits will also be assigned to teaching skills. This includes taking pedagogical courses, as well as supervising Bachelor's and Master's students at the local node.
SMARTHEP credits system:

Table 1.2.2: Example SMARTHEP doctoral program

Type of training	Number of credits
Training through research	**135**
At host	75
Through secondment	60
Training through lectures, courses and dissemination	**45**
Technical and Research Training	30
Transferable Skills Training	15
of which towards teaching and dissemination	5
Total	180

To ensure that such a diverse training program is coherent and recognized across the network, we have designed a SMARTHEP credit system according to the ECTS standard. Each ESR will complete 180 SMARTHEP credits, as shown for an example ESR with two secondments in Table 1.2.2. As PhDs in all participating institutions are awarded based on local regulations, SMARTHEP credits also ensure that ESRs receive the appropriate training. Each network-wide event below includes the amount of assigned SMARTHEP credits.

We have assigned 1 SMARTHEP credit per each 1/2 day of lectures. All students will have to explicitly include 15 credits of transferrable skills training within their PCDP - they will be able to choose among the programs of their local institute, or the institutes / industries they are seconded in. By attending the yearly meetings of the network and the final conference, around 1/2 of the SMARTHEP credits that they will need to complete in both "Technical and Research" and "Transferable Skills" categories, as presented at the beginning of the section, will be provided in network-wide events. ESRs will have freedom how to complete the rest of required credits through the local resources. Finally, ESRs will be required to present their work to at least one conference outside the network in their area of expertise. A list of conferences of interest for SMARTHEP topics is given in Sec. ??. The amount of credits awarded will depend on the targeted conference. The attribution of ECTS credits will require institutes to explicitly include the network events in their course plan, but the conversion from SMARTHEP to ECTS credits will be justified and straightforward.

B1.2.2 Role of non-academic sector in the training program

Basic research is at the heart of its PhD projects in the ESRs, but SMARTHEP dynamically and directly addresses the challenges of the academic, industrial, and entrepreneurial sectors. This enriches all sectors through the transfer of best practice, knowledge, and expertise. For this reason we have included a comprehensive program of non-academic training in SMARTHEP including hands-on experience in solving practical problems during secondments at industrial partners, discussed further in Sec. 1.3.3. The non–academic partners of SMARTHEP will have the following crucial roles in the training of the ESRs:

- **Training through research: secondments.** One of SMARTHEP's most important objectives is to increase the exposure of the students to the private sector, solving practical problems with RTA and creating commercial value. Most ESRs will have secondments at private companies relevant for their tools and research topics. These will place a particular emphasis on common methods between the commercial applications which the non–academic