MARIE Skłodowska-CURIE ACTIONS

Innovative Training Networks (ITN)
Call: H2020-MSCA-ITN-2020

PART B

“DESCARTES”

aDvancEd StatiStiCal Algorithms for paRTicLE phySics

This proposal is to be evaluated as:

[ETN]
# TABLE OF CONTENTS

1. Excellence.......................................................................................................................... 5
   1.1 Quality, innovative aspects and credibility of the research programme............................... 5
      1.1.1 Introduction, objectives and overview of the research programme..................................... 5
      1.1.2 Research methodology and approach................................................................................. 6
      1.1.3 Originality and innovative aspects of the research programme........................................... 7
   1.2 Quality and innovative aspects of the training programme..................................................... 8
      1.2.1 Overview and content structure of the training................................................................. 8
      1.2.2 Role of non-academic sector in the training programme.................................................. 10
   1.3 Quality of the supervision..................................................................................................... 11
      1.3.1 Qualifications and supervision experience of supervisors.................................................. 11
      1.3.2 Quality of the joint supervision arrangements................................................................. 13
   1.4 Quality of the proposed interaction between the participating organisations........................... 13
      1.4.1 Contribution of all participating organisations to the research and training programme......... 13
      1.4.2 Synergies between participating organisations............................................................... 14
      1.4.3 Exposure of recruited researchers to different (research) environments, and the complementarity thereof................................................................................................. 14

2. Impact.................................................................................................................................... 15
   2.1 Enhancing the career perspectives and employability of researchers and contribution to their skills development........................................................................................................ 15
   2.2 Contribution to structuring doctoral/early-stage research training at the European level........... 16
   2.3 Quality of the proposed measures to exploit and disseminate the results............................... 17
      2.3.1 Dissemination of the research results.................................................................................. 17
      2.3.2 Exploitation of results and intellectual property............................................................. 18
   2.4 Quality of the proposed measures to communicate the activities to different target audiences...... 18
      2.4.1 Communication and public engagement strategy............................................................. 18

3. Quality and Efficiency of the Implementation....................................................................... 19
   3.1 Coherence and effectiveness of the work plan....................................................................... 19
   3.2 Appropriateness of the management structures and procedures............................................. 28
      3.2.1 Network organisation and management structure......................................................... 28
      3.2.2 Supervisory board............................................................................................................ 29
      3.2.3 Recruitment strategy......................................................................................................... 29
      3.2.4 Progress monitoring and evaluation of individual projects.............................................. 30
      3.2.5 Risk management............................................................................................................ 30
      3.2.6 Intellectual Property Rights (IPR).................................................................................. 31
      3.2.7 Gender aspects................................................................................................................ 32
      3.2.8 Data management plan................................................................................................... 32
   3.3 Appropriateness of the infrastructure of the participating organisations..................................... 32
   3.4 Competences, experience and complementarity of the participating organisations and their commitment to the programme......................................................................................... 33
      3.4.1 Consortium composition and exploitation of participating organisations' complementarities........ 33
      3.4.2 Commitment of beneficiaries and partner organisations to the programme....................... 33
   3.5 Gantt Chart.......................................................................................................................... 34

4. Participating Organisations..................................................................................................... 35

5. Ethics Issues............................................................................................................................. 51

6. Letters of Commitment.......................................................................................................... 51
# LIST OF PARTICIPATING ORGANISATIONS

The following table lists all academic beneficiaries and partners of the consortium. Note: there are no inter-relationships between institutions or individuals participating in the programme.

<table>
<thead>
<tr>
<th>Consortium Member</th>
<th>Legal Entity Short Name</th>
<th>Academic (tick)</th>
<th>Non-academic (tick)</th>
<th>Awards Doctoral Degrees (tick)</th>
<th>Country</th>
<th>Dept./Division / Laboratory</th>
<th>Scientist-in-Charge</th>
<th>Role of Partner Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beneficiaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Institute of Accelerating Systems and Applications, University of Athens</td>
<td>IASA</td>
<td>X</td>
<td>X</td>
<td>Greece</td>
<td>Konstantinos Vellidis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Centre Européenne pour la Recherche Nucléaire</td>
<td>CERN</td>
<td>X</td>
<td></td>
<td>Switzerland</td>
<td>Gianluca Cerminara</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. University of Oxford</td>
<td>UOXF</td>
<td>X</td>
<td>X</td>
<td>UK</td>
<td>Dept. of Physics</td>
<td>Daniela Bortoletto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Università degli Studi di Padova</td>
<td>UNIPD</td>
<td>X</td>
<td>X</td>
<td>Italy</td>
<td>Dept. of Stat. Sciences</td>
<td>Giovanna Menardi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Université Catholique de Louvain</td>
<td>UCL</td>
<td>X</td>
<td>X</td>
<td>Belgium</td>
<td>Centre de Cosmologie, Physique des Particules et Phénoménologie</td>
<td>Christophe Delaere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Université Clermont Auvergne</td>
<td>UCA</td>
<td>X</td>
<td>X</td>
<td>France</td>
<td>LPC, Laboratoire de Physique de Clermont</td>
<td>Julien Donini</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Laboratório de Instrumentação e Física Experimental de Partículas, Lisbon</td>
<td>LIP</td>
<td>X</td>
<td>X</td>
<td>Portugal</td>
<td>LIP Lisbon</td>
<td>Michele Gallinaro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Universidad de Oviedo</td>
<td>OVD</td>
<td>X</td>
<td>X</td>
<td>Spain</td>
<td>Dept. of Physics</td>
<td>Javier Cuevas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Istituto Nazionale di Fisica Nucleare</td>
<td>INFN</td>
<td>X</td>
<td></td>
<td>Italy</td>
<td>Sezione di Padova</td>
<td>Donatella Lucchesi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. University of Oslo</td>
<td>UiO</td>
<td>X</td>
<td>X</td>
<td>Norway</td>
<td>Dept. of Mathematics</td>
<td>Riccardo De Bin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Università degli Studi di Milano Bicocca</td>
<td>UNIMIB</td>
<td>X</td>
<td></td>
<td>Italy</td>
<td>Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa</td>
<td>Nicola Lunardon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner Organisations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Location of research premises (city / country)</td>
<td>Type of R&amp;D activities</td>
<td>No. of full-time employees</td>
<td>No. of employees in R&amp;D</td>
<td>Web site</td>
<td>Annual turnover (in Euros)</td>
<td>Enterprise status (Yes/No)</td>
<td>SME status (Yes/No)</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>----------</td>
<td>---------------------------</td>
<td>----------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>12. Fermi National Accelerator Laboratory (Fermilab)</td>
<td>X</td>
<td>US</td>
<td>Sergio Jindariani</td>
<td>Research and training in ML algorithm implementation on electronic boards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. University of California, Irvine</td>
<td>X</td>
<td>US</td>
<td>Daniele Whiteson</td>
<td>Training in ML applications for physical sciences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. University of Liège</td>
<td>X</td>
<td>Belgium</td>
<td>Gilles Louppe</td>
<td>Training in deep generative models and probabilistic programming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. AI4B Limited</td>
<td>AI4B</td>
<td>X</td>
<td>Michael Spannowsky</td>
<td>Training in business development services using novel data analysis methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Achilles Information Limited</td>
<td>ACH</td>
<td>X</td>
<td>Fabrizio Margaroli</td>
<td>Training in ML algorithms for commercial and public usage applications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. TWave S.L.</td>
<td>TW</td>
<td>X</td>
<td>Juan Menéndez</td>
<td>Training in development of highly integrated instrumentation equipment and ML applications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Artabro Tech</td>
<td>ART</td>
<td>X</td>
<td>Miguel Vidal</td>
<td>Training in real data analysis and rating algorithms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following table lists all non-academic beneficiaries of the consortium.
1. Excellence

1.1 Quality, innovative aspects and credibility of the research programme

1.1.1 Introduction, objectives and overview of the research programme

Since the discovery of the Higgs boson in 2012 at the CERN Large Hadron Collider (LHC), the landscape for research of new phenomena in particle physics has changed considerably. So far there is no clear sign of “New Physics” beyond the standard model (SM), the theory that has described particle physics with unquestionable success for more than 40 years. Apart from a few deviations observed in B meson decays, that raise questions about the universality of lepton flavours couplings, all predictions of the standard model have been verified with very high precision at the LHC and other particle physics experiments. The situation is paradoxical because many theories propose to extend the standard model in order to solve its theoretical problems, such as naturalness in electroweak interaction (Higgs boson mass) or in strong interactions (CP problem in QCD), or to address experimental evidence (neutrino mass, evidence for dark matter).

This situation requires a rethinking of the strategies to unveil new physics at the LHC, also in preparation for the high luminosity phase of the LHC that will start in 2027 (HL-LHC). In addition to “traditional” searches for new resonances, leading to clear new physics signatures, more global approaches are needed. For example, the current efforts to interpret LHC measurements in an effective field theoretical framework (such as SM-EFT) are in line with this goal. A complementary approach is to probe, using high-performance statistical methods, a large space of parameters, without a priori on the nature of the researched signal. The Higgs boson also plays a fundamental structural role in the standard model theory, and measuring its properties with the highest possible accuracy may shed light on new phenomena.

All these approaches benefit greatly from modern data science methods, as advances on algorithmic, software and hardware techniques enable to develop fast and robust solutions for a large variety of use-cases in the high energy physics (HEP) field. The widespread usage of machine learning (ML) techniques in the HEP community is now remarkable, in particular given that not so long ago multivariate methods were relying mostly on “traditional” techniques (such as boosted decision trees or shallow neural networks), and the continuous advances in data science can make the HEP field benefit from the application of unexplored, state of the art techniques.

The DESCARTES project brings together particle physicists and statisticians to improve data science techniques and to extend our understanding of fundamental physics. The main objective is to exploit LHC data (and future HL-LHC data) to extend our knowledge in two of the most important domains in particle physics nowadays: directly searching for potential new physics signatures and completing our understanding of the nature of the Higgs boson. The heart of our research programme relies on the following aspects:

- **Scientific innovation**: we aim at developing advanced methods for data analysis (searches, measurements), data optimization (reconstruction, identification), and data selection (triggering) for experimental particle physics. The project naturally fosters innovative developments from its proximity to similar statistical aspects present in other fields (e.g. medical physics imaging, biosciences, finance) and in industry use-cases.

- **Technical innovation**: DESCARTES is expected to significantly improve the performances of HEP analyses by exploiting recent developments in computer science (popular algorithms such as Generative Adversarial Networks only date back to 2014) and technological progresses enabling fast inference (using FPGA) and high performance large scale processing (with e.g. Apache Spark).

- **Collaboration with Computer Scientists and statisticians**: these communities have great scientific and technical expertise in the field of data science. Collaboration with partners in these fields allows physicists to be quickly informed of potentially interesting algorithmic developments and also to benefit from fast and efficient practical implementations.

- **Collaboration with industry**: Collaboration of academic and industry partners active in the data science field allows for a mutually beneficial exchange of experience, skills, and knowledge.

- **Training**: knowledge acquired in the field of statistics, machine learning, data analysis is of interest to a wide scientific community and a demand for training from students (Master, Ph.D.) and researchers has been clearly identified. DESCARTES will make sure to incorporate these aspects while conducting the research programme.
1.1.2 Research methodology and approach

The core of the DESCARTES scientific program is devoted to research in experimental particle physics, with several laboratories and universities participating in the CERN experiments and at the LHC. The network fosters collaboration and connection between statisticians, phenomenologists, and industrial partners in the development of methodological and technological aspects of the research program.

Data Science constitutes the backbone of the research methodology. In fact, particle physics is an ideal test bed for the development and application of modern statistical and ML techniques, made efficient by smart computing environments: the amount of available experimental data is very large, accurate Monte Carlo simulations provide extensive samples of labelled data to train ML algorithms, researched signals are rare and complex thus requiring advanced statistical approaches to separate them from an overwhelming background, extreme data taking conditions need fast and efficient online algorithms and hardware, etc.

Modern ML approaches are at use in a large variety of particle physics use-cases: regression (e.g. data modelling, physics object reconstruction and resolution), classification (e.g. particle identification, signal separation), generation (fast detector simulation), unsupervised learning (e.g. clustering, anomaly detection), etc. Common algorithms include deep MLP networks, but also more advanced architectures such as CNN, GAN, VAE or LSTM are being used. Advanced machine learning libraries (such as Tensorflow) can now be interfaced with the complex software frameworks used by LHC collaborations and can be part of the production workflow. On the other hand, some of these methods are still in the R&D phase and have not reached yet an optimal efficiency and some data science approaches are still unexplored in the context of HEP (among others, methods for network or functional data as well as complex pseudo-likelihoods); more generally new ML approaches and developments are expected to improve the overall performances in many areas. These ML methods, algorithms and tools are also of interest for other scientific fields and the industry, thus enabling the possibility of fruitful collaborations.

The research program of DESCARTES is structured in four scientific work packages (WP), each addressing a specific research area. These WP are connected and share expertise, methods and results.

- **WP1**: Searches for new physics in LHC data. Searches are performed in two main directions: model independent general searches, where large sectors of phase space are probed to search for deviation from standard model predictions; and searches for long-lived particles that are predicted by several theoretical models and which constitute a challenging and original approach to look for new phenomena. These searches rely on semi-supervised advanced models for anomaly detection, as e.g. autoencoders, based on neural network structures.

- **WP2**: Measurements of Higgs boson properties with LHC data. Understanding the fundamental nature of the Higgs boson is central to the SM and also to theories beyond the SM: If the Higgs mechanism was found to be non-SM like this would have huge theoretical repercussions and would certainly point towards the existence of new interactions and particles. In this WP several ML approaches are going to be proposed to enhance the identification of rare Higgs signatures (for different production and decay modes) with respect to large backgrounds. Novel measurements of the Higgs boson cross-section will be performed in channels that have never been observed before. In addition, prospects for the discovery of di-Higgs (HH) production will be studied in the context of HL-LHC.

- **WP3**: Effective field theories (EFT) for physics at the LHC. The EFT framework is strongly motivated by the theory community, the paradigm being that New Physics (NP) could be located at an energy scale much higher than the energy available in LHC collisions, and therefore out of direct reach. However, carefully constraining the parameters of the EFT using global fits to a large set of measurements, would allow being indirectly sensitive to the presence of NP, thus providing a direction where to invest research efforts (for example, in the building of future particle accelerators). This WP mainly focuses on the development of novel statistical methods (deep learning-based algorithms, pseudo-likelihood functions) for constraining EFT parameters in the Higgs boson sector. In addition, this project will implement distributed computing tools for fast numerical calculations that will be beneficial to other WPs.

- **WP4**: Real-time algorithms for data selection and medical imaging. This WP aims at developing fast online data selection systems to exploit the high collision rate (40 MHz) and high data volume (hundreds Pb/year of stored data) produced at the HL-LHC. To this end ML methods will be run for fast inference on programmable hardware (such as FPGA integrated circuits). The technologies and methodologies that will be developed within this WP are potentially interesting for several LHC experiments and a transfer of knowledge could be performed through the WP 1-3. Another original aspect of this WP is that it will be conducted in synergy with projects on medical imaging where similar real-time reconstruction and classification ML algorithms can be efficiently used.
1.1.3 Originality and innovative aspects of the research programme

DESCARTES ambitious project covers a broad range of scientific developments focused on the four WP objectives described in Section 1.1.2. The originality of the DESCARTES approach revolves around the collaboration of experimental physicists, phenomenologists, and statisticians. All WPs are devised to profit at the best of this interaction, for example, involving the statisticians to provide guidance and expertise in the algorithm development phase of the research. Furthermore, the network is designed to maximize the exchange between the academic and industrial sectors for what concerns innovative statistical algorithms, analysis methodologies, and technologies. The fulfillment of these goals requires an extensive network, including academic and industrial beneficiaries and partners, providing the 14 ESR with a vast portfolio of expertise in different disciplines and sectors.

Table 1.1: Work Package (WP) list

<table>
<thead>
<tr>
<th>WP No.</th>
<th>WP Title</th>
<th>Lead Beneficiary No.</th>
<th>Start Month</th>
<th>End Month</th>
<th>Activity Type</th>
<th>Lead Beneficiary Short Name</th>
<th>ESR involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Searches for new physics in LHC data</td>
<td>6</td>
<td>1</td>
<td>48</td>
<td>Research, interdisciplinary</td>
<td>UCA</td>
<td>1,3,4,6,7,9</td>
</tr>
<tr>
<td>2</td>
<td>Measurements of Higgs boson properties with LHC data</td>
<td>9</td>
<td>1</td>
<td>48</td>
<td>Research, interdisciplinary</td>
<td>INFN</td>
<td>3,4,6,7,8,10,12</td>
</tr>
<tr>
<td>3</td>
<td>Effective field theories for physics at the LHC</td>
<td>3</td>
<td>1</td>
<td>48</td>
<td>Research, interdisciplinary</td>
<td>UOXF</td>
<td>5,8,9,10,11,12</td>
</tr>
<tr>
<td>4</td>
<td>Real-time algorithms for data selection and medical imaging</td>
<td>2</td>
<td>1</td>
<td>48</td>
<td>Research, interdisciplinary</td>
<td>CERN</td>
<td>2,4,13,14</td>
</tr>
<tr>
<td>5</td>
<td>Training and career development</td>
<td>5</td>
<td>1</td>
<td>48</td>
<td>Career planning, interdisciplinary</td>
<td>UCL</td>
<td>All</td>
</tr>
<tr>
<td>6</td>
<td>Dissemination and outreach activities</td>
<td>7</td>
<td>1</td>
<td>48</td>
<td>Public engagement and dissemination</td>
<td>LIP</td>
<td>All</td>
</tr>
<tr>
<td>7</td>
<td>Management activities</td>
<td>1</td>
<td>1</td>
<td>48</td>
<td>Management</td>
<td>IASA</td>
<td>ESR in S. B.</td>
</tr>
</tbody>
</table>

Our project builds on the extensive experience that the experimental particle physics community has now achieved in modern machine learning techniques. Several of DESCARTES members are actively participating in statistical and machine learning forums and workshops at CERN and elsewhere. All this provides a solid ground on which DESCARTES will develop new machine learning approaches that will be applied to all areas of the research programme.

Besides, each of the scientific WP will promote developments of cutting-edge scientific field and technological applications, namely:

1. Modern data science approaches for anomaly detection (WP1). This theme is of interest well beyond physics and calls for a vast variety of industrial use-cases (fraud detection, etc).
2. Data processing with fast and scalable ML tools using the Apache spark ecosystem (WP1-3). The data produced for LHC analyses are an excellent opportunity for the development and application of "big data" analysis paradigms. These are of interest to many other application fields, including the private sectors.

3. Deployment in FPGA electronics of ML algorithms for fast and precise reconstruction, identification and selection (WP4). The interest in FPGA hardware is growing for a wide variety of application fields, progress in the deployment of complex algorithms in firmware can further expand the potential of these devices. This task will build on CERN experience in the development of High-Level Synthesis for Machine Learning tools.¹

1.2 Quality and innovative aspects of the training programme

1.2.1 Overview and content structure of the training

The network is designed to give to the early stage researchers the best assets to do excellent research in particle physics with solid statistical foundations. The strong connection with the industry both aims at giving them additional competencies that they can use in fundamental research and at making them aware of the transferable skills acquired that they can value in their future professional life. The skills that they will acquire can therefore be divided in three categories:

Core Research Skills. These will be acquired through the individual research projects. They concern particle physics, data analysis, data modelling, and use of advanced ML tools. These are areas where the ESRs are expected to become experts. Additionally, they are expected to develop key competences for lifelong learning, including confident and critical use of information and communications technology, intellectual freedom, professional responsibility and accountability.

Advanced/Additional Skills. These will be offered by the consortium. The research program will be conducted in a variety of academic and industrial sites across Europe, forming together an attractive institutional environment for the ESRs. This environment will provide the ESRs with interdisciplinary research experience and exposure to relevant employment sectors. This will be achieved by a number of secondments within the network, designed to consolidate the synergy of research groups in relevant activities, as well as to offer the ESRs the experience of working on topics different than those of their individual research program. In addition, visits to CERN are foreseen for the ESRs involved in CERN experiments, to meet with larger groups relevant to their projects and participate to on-site experimental activities, as well as to conduct the services required to gain the credit of co-signatory author in publications from the respective Collaborations. The ESRs are expected to develop the ability to manage their own learning, either individually or in groups, and to participate effectively and constructively in societal and working life.

Transferable Skills. These will also be delivered by the consortium. The training program envisions a number of network meetings as well as training, outreach, and dissemination events, including seminars, schools and open workshops. Furthermore, the program includes participation in international conferences and workshops that will help the ESRs develop their communication skills by presenting their research results in talks or posters. The ESRs are expected to acquire excellent written and oral communication and presentation skills, communication and dialogue with non-technical audiences, dissemination, public engagement, and teaching skills.

Three modes of training delivery are envisioned:

- **Local training**: this training will be delivered to the ESRs at their main host institution in terms of their individual research programs and graduate program. It concerns research-related training, developing skills such as knowledge of research methodologies and technologies, research ethics and integrity, and good practice in research (e.g. reliable backing up of data). Local training will be enlarged through the network, where training available in some member institution is opened up to ESRs from the other institutions in the consortium, for example by visits to CERN.

- **Network-wide training**: this mode aims at imparting complementary skills, related with communication, dissemination, outreach and career development skills. The training will be delivered through a number of network events, including some opened up to the wider research community. See Table 1.2b. We ensured to propose a variety of events covering all the three types of skills. It goes from very topical training in particle physics (e.g. MadGraph) to transversal scientific statistical tools, and is complemented with soft skills (e.g. Scientific Communication).

- **Secondment program**: a number of secondments are foreseen, see Section 1.4.

¹ HLS4ML: https://fastmachinelearning.org/hls4ml/
A Personal Career Development Plan (PCDP) will be prepared between the ESR and the corresponding Supervisory Committee in order to properly balance the local and network-wide training programs. This deliverable will include:

- A personalised analysis of the requirements and goals of the planned training for the ESR.
- A list of courses (local and network-wide) to be taken by the ESR during the training program, including any ECTS credit requirements.
- A list of communication and dissemination activities to be undertaken by the ESR.
- A schedule for the ESR’s program, including secondments.

**Table 1.2a: Recruitment deliverables per beneficiary**

<table>
<thead>
<tr>
<th>Researcher No.</th>
<th>Recruiting Participant (short name)</th>
<th>PhD awarding entities</th>
<th>Planned Start Month 0-45</th>
<th>Duration (months) 3-36</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>IASA</td>
<td>National and Kapodistrian University of Athens</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>2.</td>
<td>UCA</td>
<td>Université Clermont Auvergne</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>3.</td>
<td>UNIPD</td>
<td>University of Padova, Dept. of Stat. Sciences</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>4.</td>
<td>INFN</td>
<td>University of Padova, Dept. of Physics and Astronomy</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>5.</td>
<td>LIP</td>
<td>University of Lisbon</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>6.</td>
<td>OVD</td>
<td>University of Oviedo</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>7.</td>
<td>UiO</td>
<td>University of Oslo</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>8.</td>
<td>UOXF</td>
<td>University of Oxford</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>9.</td>
<td>UCL</td>
<td>Université Catholique de Louvain</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>10.</td>
<td>UNIMIB</td>
<td>University of Milano Bicocca</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>11.</td>
<td>B12</td>
<td>Université Catholique de Louvain</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>12.</td>
<td>CERN</td>
<td>University of Oviedo</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>13.</td>
<td>IASA</td>
<td>National and Kapodistrian University of Athens</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>14.</td>
<td>KROMEK</td>
<td>University of Durham</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>504</td>
</tr>
</tbody>
</table>

**Table 1.2b: Main network-wide training events, conferences and contribution of beneficiaries**

<table>
<thead>
<tr>
<th>Main Training Events &amp; Conferences</th>
<th>ECTS (if any)</th>
<th>Lead Institution</th>
<th>Action Month (estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Network kick-off meeting.</td>
<td></td>
<td>IASA</td>
<td>4</td>
</tr>
<tr>
<td>2 1st network meeting + school of statistics and statistical computing tools offering a common background on these topics.</td>
<td></td>
<td>UNIMIB</td>
<td>10</td>
</tr>
<tr>
<td>3 MadGraph school offering background knowledge on the use of matrix elements in effective field theories.</td>
<td></td>
<td>UCL</td>
<td>12</td>
</tr>
<tr>
<td>4 2nd network meeting + outreach workshop presenting the network objectives and plans to interdisciplinary academic and high school students.</td>
<td></td>
<td>LIP</td>
<td>16</td>
</tr>
<tr>
<td>5 School on ML methods for big data, including examples from applications of these methods in physics.</td>
<td></td>
<td>UiO</td>
<td>19</td>
</tr>
<tr>
<td>6 School of statistics organised in France, offering an overview of the statistical concepts and machine learning tools used in particle physics.</td>
<td></td>
<td>UCA</td>
<td>22</td>
</tr>
<tr>
<td>7 3rd network (mid-term) meeting + outreach “hackathon”.</td>
<td></td>
<td>CERN</td>
<td>24</td>
</tr>
<tr>
<td>8 Science communication workshop.</td>
<td></td>
<td>UOXF</td>
<td>27</td>
</tr>
<tr>
<td>9 Computing technology workshop on ML algorithm implementations on hardware.</td>
<td></td>
<td>FNAL</td>
<td>30</td>
</tr>
<tr>
<td>10 4th network meeting + knowledge transfer workshop on machine learning applications in cosmic ray tomography.</td>
<td></td>
<td>UCL</td>
<td>33</td>
</tr>
<tr>
<td>11 Soft skills workshop</td>
<td></td>
<td>KROMEK/B12</td>
<td>36</td>
</tr>
<tr>
<td>12 5th network meeting + thematic workshop for dissemination in industrial sector co-organised with the CERN Knowledge-Transfer group.</td>
<td></td>
<td>CERN</td>
<td>40</td>
</tr>
<tr>
<td>13 Knowledge transfer workshop involving statisticians working on physics and physicists working on statistics.</td>
<td></td>
<td>UNIPD</td>
<td>41</td>
</tr>
</tbody>
</table>
1.2.2 Role of non-academic sector in the training programme

Industrial partners as beneficiaries, with the responsibility of shaping a joint research program relevant to common activities with academic participants of the network, and recruiting and co-supervising ESRs

We envision the role of non-academic partners in DESCARTES to serve as a catalyst for bridging the gap between academia and industry/business. One part of our vision is to contribute to the continuous improvement of availability and quality of highly skilled through synergy of academia and industry. Another is to allow for faster adoption of novel machine/deep learning, big data and project management techniques in HEP, with the ultimate aim of advancing scientific research.

B12 Consulting

As a beneficiary, B12 consulting will host a PhD student in close collaboration with (and under the mentorship of) Andrea Giammanco from the Center for Cosmology, Particle Physics Phenomenology (CP3) at UCL. Geographic proximity of B12 to the CP3 institute, as well as the excellent existing cooperation between the two entities (e.g. past collaboration with AMVA4NewPhysics and MCNET) will allow ESRs to maximally benefit from the simultaneous exposure to non-academic and academic environments: students can easily participate in academic activities of CP3 (courses, seminars, research collaborations...) as well as maintain activity in business related activities of B12 (commercial projects). B12 features a portfolio counting nearly 100 projects in domains of automotive industry, mobility, pharmaceuticals, health, social networking, aerospace, finance and education, allowing ESRs to have exposure to a vast palette of industry and business domains. Technological expertise of B12 includes (but is not limited to): deep learning, classical machine learning, advanced statistical techniques, commercial software development, database management and IT strategy consulting, perfectly in line with the technologies and objectives of DESCARTES.

ESRs placed at B12 will work on a HEP related PhD project while absorbing techniques and experiences from the private sector which are not a common part of the academic curriculum. B12 plans to emphasize strong coding skills and practices, expertise in modern deep learning technologies as well as development of solid agile management skills, communication skills and business acumen. On the academic side, B12 will co-mentor the ESR’s HEP project, evangelise the use of above-mentioned techniques within HEP, and work to improve the adoption of new technologies developed by the DESCARTES collaboration. Other partners will host the secondments of ESRs and provide their complementary expertise in applications beyond the scope of particle physics.

KROMEK Ltd

KROMEK specialises in the development and production of radiation detector components and end-user products. These include X-ray, gamma ray and neutron detection products for the security, nuclear and health markets. As a beneficiary KROMEK will host one PhD student who will receive training in data analysis and machine learning techniques which can be applied directly to real-life problems. While being placed at KROMEK, the ESR will work on a project related to particle detection phenomenology, as outlined in the grant proposal, but will have the opportunity to be involved in commercial projects as well.

Having span out of the Physics Department in 2003, KROMEK maintains numerous links with the University of Durham. Several students from the Physics Department have carried out internships at KROMEK, and have subsequently become permanent employees at the company. The ESR will be enrolled as a PhD student at the University, with Prof. Michael Spannowsky acting as the academic supervisor. This will allow the ESR to attend training courses, workshops and lectures offered by the University. The ESR will also carry out a secondment at the University’s Institute for Particle Physics Phenomenology (IPPP). The IPPP is the largest university-based institute for particle physics phenomenology in the world and has a strong research focus on machine learning and data analysis techniques.

Artabro Tech

ART will participate in the network by hosting the secondments of two ESRs and provide their expertise in the application of AI to non-academic problems. The main activities of the company are related to Industry 4.0 and Business intelligence services. In the case of Industry 4.0, the ESRs will have a picture of the whole data cycle related to a production plant (data collection from industrial PLCs, digital twin architecture and plant control). The ESR will acquire experience in training and deploying Neural Networks to predict industrial set points in a real production line. In the case of Business Intelligence the ESRs will be in touch with data related to business
opportunities, tender data from different countries and private business opportunities. The ESR will acquire capabilities in real data analysis and rating algorithms.

**TWave S.L.**

TW will participate in the network by hosting the secondments of two ESRs and offering expertise in the technologies and tools related with the company’s business activity, which is the development of highly integrated instrumentation equipment. During their secondments, specific tasks faced by the ESRs will involve ML applications on signal classification and fault detection.

**Achilles Information Ltd**

ACH will participate in the network by hosting the secondments of two ESRs and providing its leading expertise in the application of innovative statistical and machine learning tools for near-real time risk management of supply chain risks. During their secondments the ESRs will analyse data from suppliers world-wide and will receive training in advanced machine learning tasks and techniques for reading vast volumes of data from relational, non-relational and graph databases.

**AI4B**

AI4B will participate in the network by hosting the secondment of one ESR and by providing leading expertise in the application of unsupervised and supervised machine learning tools to perform classification and regression tasks. During the secondment the ESR will work as part of a team to address data-science-intensive tasks which the client needs to solve to improve its company’s performance. This includes the consultation of the client and the implementation of software solutions.

### 1.3 Quality of the supervision

#### 1.3.1 Qualifications and supervision experience of supervisors

Most of the supervisors are academics with permanent Professor positions. They all have a well-established supervision experience with several PhD theses completed under their supervision, as summarized in Table 1.3.

**Table 1.3: Qualifications and supervision experience of main supervisors in beneficiary nodes**

<p>| Konstantinos Vellidis | Vellidis (Ph.D. 2001) is Associate Professor at the Dept. of Physics of the National and Kapodistrian University of Athens (NKUA) since 2017. His PhD is in nuclear physics (search for signatures of deformed baryon states). 4-year participation in electron scattering experiments at the Bates Accelerator of MIT. He moved to particle physics in 2004, joining first the ATLAS and then the CDF experiments. He worked at Fermilab (2008-2017) in the CDF and Mu2e experiments and at CERN (2017-2018) in CMS. As a Fermilab scientist, he mentored and co-supervised 2 post-docs, 3 PhD students and 6 MSc students, and co-organised summer internships for graduate students for four years (2013-2016). Currently he is main supervisor of one PhD student and co-supervisor of two more PhD students at NKUA. |
| Julien Donini | Donini (Ph.D. 2002) is Professor at the University Clermont Auvergne (UC) where he teaches courses of Physics, Particle Physics, Statistics and Machine learning. In 2017 he created, and currently manages, a new Data Scientist degree at the university. Since 2012, he has been a member of the organizing committee of the IN2P3 School of Statistics, organized every two years in France. In the last 5 years he supervised 3 post-docs, 5 Ph. D. students and 6 Master students. Donini is member of the ATLAS collaboration since 2008 and his research activities focuses on top-quark physics and searches for new physics. |
| Giovanna Menardi | Menardi (Ph.D. 2007) is Associate Professor at the Department of Statistical Science, University of Padova. After obtaining a Ph.D. in Statistics Applied to Social Sciences and Economics in Padova, she worked as a research fellow at the Universities of Trieste and Padova and as Assistant Professor at University of Padova. She has taught graduate courses at the University of Trieste, Padova and Ca’ Foscari (Venezia) and has given post-graduate lectures in Masters in E-business and in Spatial Geographic Information System (University of Trieste) and in Training advanced courses for business (Generali Group). She has supervised several undergraduates and master students, two PhD students and a post-doc. Her research activity has been presented in many international conferences and published in journals. |
| Donatella Lucchesi | Lucchesi (Ph.D. 1994) is full Professor at the Dept. of Physics and Astronomy, University of Padova and has a research appointment at INFN. She served as Software and Computing coordinator at CDF during the transition to the GRID paradigm. She contributed to the $B_s$ oscillation frequency measurement and in the last period of CDF she has been one of the most active members of the Higgs physics group supervising three students. She is now, member of LHCb, in charge of the LHCb Padova unit, and she is contributing to Higgs and new resonance searches in b-jet finals states in the forward region covered by LHCb. She supervised two PhD students on these measurements. |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Position</th>
<th>Department/Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michele Gallinaro</td>
<td>LIP</td>
<td>Supervisor of ESR5</td>
<td>Gallinaro (Ph.D. 1996) is professor at the University of Lisbon where he teaches courses at the master and doctorate level, and researcher at LIP (Laboratório de Instrumentação e Física Experimental de Partículas). He has long experience in Hadron Collider physics including the Tevatron and the LHC, both in detector development and data analysis. At the Tevatron he contributed to the discovery of the top quark with the CDF experiment in 1995. In the CMS collaboration, he coordinated the group physics analysis efforts during Run1. He coordinated the “Physics Performance” and “Timing detector” groups of the CMS Precision Proton Spectrometer. He coordinated the scientific work of 7 PhD students, 14 undergraduate students, and several postdoctoral fellows (10+). He is co-author of more than 1500 scientific publications (90 average citations per paper).</td>
</tr>
<tr>
<td>Javier Cuevas</td>
<td>OVD</td>
<td>Supervisor of ESR6</td>
<td>Cuevas (Ph.D. 1987) is full professor at the Department of Physics, University of Oviedo since 1994. Where he has been teaching advanced and introductory classes for undergraduate students. Head of the HEP group at University of Oviedo. Member of the DELPHI experiment at LEP, CDF at Tevatron and CMS at LHC. Deputy manager of the Spanish HEP program of the spanish agency for research in the periods 2014-2016 and from June 2018 to date. He has supervised over 9 Ph.D. students and several post-docs and master students. He has directly participated in more than 100 publications related with Higgs physics, SM measurements, top-quark physics and BS M searches.</td>
</tr>
<tr>
<td>Riccardo De Bin</td>
<td>UiO</td>
<td>Supervisor of ESR7</td>
<td>De Bin (Ph.D. 2012), is Associate Professor at the Department of Mathematics of the University of Oslo since 2017. Before he was Assistant Professor at the Radboud University Medical Centre of Nijmegen (The Netherlands, 2016), post-doctoral fellow at the Lüdwg-Maximilians-Universität of Munich (Germany, 2012-2016) and pre-doctoral fellow at the Northwestern University of Evanston, USA (2010-2011). He taught at BSc, MSc and PhD level in 4 different countries. He supervised (including students currently enrolled) 3 PhD students as main supervisor, 3 PhD students as co-supervisor, 11 MSc students and 4 BSc students.</td>
</tr>
<tr>
<td>Daniela Bortoletto</td>
<td>UO XF</td>
<td>Supervisor of ESR8</td>
<td>Bortoletto (Ph.D. 1989) was the E. M. Purcell distinguished professor of Physics at Purdue University. She is currently the head of the sub department of Particle Physics at the University of Oxford. For two decades Prof. Bortoletto has taught advanced and introductory classes receiving excellent evaluations from the students and the Spira teaching award at Purdue University. She has supervised 18 Ph. D. students, 12 post-docs, and mentored over 35 undergraduates in research. Her students and postdocs are now faculty members at major universities all over the world. She is a sought after speaker at international conferences and taught at many summer schools. Bortoletto is passionate about gender issues and increasing female participation in physics. She has been the first chair of the National Organizing Committee of the APS Conference Undergraduate Women in Physics (CUWiP). She has brought CUWiP to the UK where the conference has been extraordinarily successful.</td>
</tr>
<tr>
<td>Christophe Delaere</td>
<td>UCL</td>
<td>Supervisor of ESR9</td>
<td>Delaere (Ph.D. 2005) has been FNS Research Associate at UCL since 2009, after spending a few years at CERN as fellow or under other postdoctoral contracts. He is now Senior Research Associate and Professor. He supervised four completed PhD theses so far, among which one was in the context of a previous ITN project, and four others are ongoing. He also supervised seven postdocs, three of who have now a tenure or tenure-track position. Member of the CMS and ALEPH collaborations at CERN and responsible for the Delphes and MoMEMta projects; his research interests range from the Higgs sector of the Standard Model to Machine Learning and Data Science.</td>
</tr>
<tr>
<td>Nicola Lunardon</td>
<td></td>
<td>Supervisor of ESR10</td>
<td>Lunardon (Ph.D. 2012) is associate professor at the Department of Economics, Quantitative Methods and Business Strategy, University of Milano Bicocca since 2019 where he was formerly research assistant since 2016. He supervised several undergraduate and graduate students and has been instructor of PhD courses at University of Milano Bicocca and at Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, US. His research activity, ranging from likelihood-based inference for complex data structure to computing methods for statistics, has been published on the top journals on statistics.</td>
</tr>
<tr>
<td>Mihailo Backovic</td>
<td></td>
<td>Supervisor of ESR11</td>
<td>Backovic (PhD 2011) is a project manager at B12-Consulting. He has completed two postdoc appointments, one at the Weizmann Institute and the other at CP3-UCL, before joining B12 in 2017. He has gained valuable teaching experience through teaching assistantship in introductory level physics courses. He has worked on a number of collaborative projects while in academia, spanning both theoretical and experimental particle physics.</td>
</tr>
<tr>
<td>Gianluca Cerminara</td>
<td>CERN</td>
<td>Supervisor of ESR12</td>
<td>Cerminara (Ph.D. 2006) is Staff Physicist in the Experimental Physics department at CERN since 2013. After obtaining his Ph.D in Experimental Particle Physics at the University of Torino, he worked as post-doc for INFN and for Northeasters University in Boston and as research fellow postdoc at CERN. He presented his work in several international conferences and published on scientific journals. He supervised more than 10 undergraduate students and one postdoctoral fellow. He is currently co-supervising a Ph.D. student in computer science on machine learning applications for anomaly detection and supervising a CERN fellow postdoc working on in research physics topics complementary to those of the proposed project.</td>
</tr>
<tr>
<td>Efstathios Stiliaris</td>
<td>IASA</td>
<td>Supervisor of ESR13</td>
<td>Stiliaris (Ph.D. 1990) is Associate Professor at the Dept. of Physics of the NKUA, where he teaches introductory Physics courses and Nuclear Physics and Experimental Methods for advanced and graduate students. He also coordinates the introductory Physics Lab and organizes the experimental program, acting in parallel Head of the Nuclear Physics Lab. He has supervised 3 PhD to conclusion, 4 PhD students in course, 9 MSc students and over 70 diploma theses.</td>
</tr>
<tr>
<td>Ben Cantwell</td>
<td>Kromek</td>
<td>Supervisor of ESR14</td>
<td>Cantwell (Ph.D. 2005) is Innovation Director of Kromek, having co-founded the company to commercialise the research into the growth of radiation detector crystals at the University of Durham in 2000/03. Cantwell has carried out number roles as the company has grown from 2 people to over 125. He has lead a number of funded R&amp;D projects, including several with university collaborations.</td>
</tr>
</tbody>
</table>
To avoid duplication, the role and scientific profile of the supervisors is only listed in the "Participating Organisations" tables (see Section 5 below).

1.3.2 Quality of the joint supervision arrangements

Each ESR will be embedded in well structured research groups composed of professors, post-doctoral researchers and other graduate students. In all cases, the supervisor remains the person to whom Early-Stage Researchers can refer for the performance of their professional duties. ESRs are informed of this when recruited.

The local supervision will be complemented by partners through secondments targeting specific HEP aspects, statistics, and industrial applications. DESCARTES will adopt a model of supervision arrangements that guarantees consistency throughout the career of the ESRs (see Fig. 1).

![Joint supervision scheme. From recruitment to graduation (ESR timeline) ESRs receive physics advisory (green) from their supervisors and from the network. During the WP tasks they will be in close connection with other network members who will provide joint supervision of the ESR activities. From the start of their secondments they will be in contact with their secondment supervisors to receive advisory in statistics and ML (red) and career development skills (blue). Successive instances of the pie chart indicate in darker colour the fraction of training received by the ESRs in the course of time.](image)

The supervision agreement is defined by the following guiding principles:

**Excellent Monitoring**: The work of ESRs will be monitored by their main supervisors and the Supervisory Board. During secondments, host supervisors will become the referent and main contact for the ESR. This complementary supervision will be the seed for a continued exchange following the secondment, in the form of topical advices through the completion of the ESR's project. Homogeneity across the project is guaranteed by the Training Coordinator who supervises the training of the ESRs and provides independent advice to aid their career development plan.

**Strong interaction within each WP**: Each work package is organized as a group of closely collaborating participants. ESRs will contribute to the WP through their individual research projects and their academic secondments. They will benefit from the collaboration between participating nodes, each one bringing its own expertise. They will also collaborate with all network participants to carry out the interdisciplinary and outreach activities.

**Topical secondments**: The secondments are arranged in order to provide complementary expertise to each ESR. For example, the typical secondment plan of an ESR attached to a HEP node consists of one secondment in another HEP node with orthogonal expertise or involved in another experimental collaboration, one secondment in a node with expertise in statistics, and one secondment in the industry.

**Strong interaction with the industrial sector**: The multidisciplinary approach of the industrial sector in problem solving will offer the ESRs a valuable experience, complementing the academic educational model and providing attractive perspectives in their career development.

1.4 Quality of the proposed interaction between the participating organisations

1.4.1 Contribution of all participating organisations to the research and training programme

The WPs have been defined to allow all participants of the network to contribute to multiple areas of research. This is possible by virtue of the existence of a common denominator in those research topics: the techniques developed in WP1, WP2, and WP3 are strongly connected. The selection strategies developed for the analysis in WP1 and WP2 drive the design of the upgraded experiments for the HL-LHC and, for this reason, are strongly connected to the research work happening in WP4. Ideas and expertise will also be shared between WP1 and WP2. Finally, all participants will be involved in the interdisciplinary and outreach activities which constitute the nucleus of WP6.
Members will be encouraged, by the network structure and by the frequent network meetings, workshops, and schools, to contribute to several different activities besides those mainly taking place at their Institutions. We will establish regular periodic monthly meetings to be held by videoconference, enabling participants to provide prompt feedback to one another on all the research and training activities. The collective know-how of network participants will thus be exploited to the fullest, maximizing the quality of training and the research output.

One specific aspect of this network needs to be addressed. Experimentalists in the network belong to either ATLAS, CMS, or LHCb; experimental data are private and sharing of internal results is forbidden. WPs address this by envisioning different analyses in different network nodes; as results on real data start to be produced, network meetings will feature plenary and parallel sessions accommodating common and experiment-specific discussions, respectively. Software development instead will be fully shared; simulation studies are based on a detector-unspecific simulation named “Delphes” purposely to allow all members to work together.

1.4.2 Synergies between participating organisations

This network is built around a strong pre-existing synergy and a history of fruitful collaboration between the proponents, pulled together by their common interests in the topics on which the network research is based. The synergy is built from the fundamental elements that form the network; the capability of effectively working together has been proven already. To quote a few examples of previous collaborations, PIs of B12, IASA, LIP, OXF, and UCL collaborated in the ITN network "AMVA4NewPhysics"; B12 and UCL also collaborated in the framework of the previous ITN network "MCNET". PIs of LIP and OXF collaborated on other topics on the observation of fully-leptonic decays of top quark pairs in 1995 (add Ref) and Higgs boson studies (add Refs?).

The ESR assignment to the activities within the four WPs has been designed to maximize the interaction and the synergy between the institutions, with each node participating in at least two (and possibly three) WPs, compatibly with the possibility of the foreseen tasks. This is with the goal of optimising the overlap and interaction of the participants. To this aim, particular care has been put in defining the schedule of secondments periods, which are functional to the specific needs of the WPs and their timeline, and which are organized to allow periods during which teams of three or four trainees physically meet and work together in the same node.

Figure 2: Planned secondments in DESCARTES.

The synergy of the envisioned collaboration is evident in the following grand-summary plan. WP1, WP2, and WP3 constitute the place where studies will be directly tested on the available data. Studies of the Higgs boson properties (WP2) and searches for New Physics (WP1) will allow direct verifiability a quantifiable measurement of the improvements developed. Additional sensitivity can be reached through indirect measurements by global fits in the EFT parameters (WP3). All reconstruction and algorithms developed can then converge in WP4 for the benefit of future studies.

Strong synergy elements are present in the way the beneficiary organizations have been distributed in the various research WPs, exploiting the know-how and experience of the members of the network for the joint development of products.

1.4.3 Exposure of recruited researchers to different (research) environments, and the complementarity thereof

The network brings together top universities and research centres as well as leading companies in specific applications and developments. This creates an international forum where trainees have privileged access to top experts in their research topic. Trainees will benefit from an intense enhanced exposure and training by interacting both with experienced experimentalists and leading phenomenologists and experts in the field, as well as with statisticians and data scientists. Thus, they will develop a more diverse portfolio of skills than what PhD students in HEP are usually exposed to.

The foreseen secondment periods at non-academic nodes will offer a different and more general approach to problem solving, broadening the skills of ESRs on topics by enabling them to find innovative ideas and transfer their abilities to other fields of research and outside academia.

In addition, the partnership with private companies envisioned by the network will ensure that the ESRs will have ample opportunity to extend their basic knowledge during their secondment time at those centres. The tight interaction with co-supervisors from partner companies foreseen by the network structure will offer ESRs opportunities that are not usually available to graduate students in standard doctoral courses.
2. Impact

2.1 Enhancing the career perspectives and employability of researchers and contribution to their skills development

**Perspectives/employability**

The DESCARTES project is crafted with particular emphasis on maximising the potential for excellent employment opportunities of our graduates both within an academic curriculum and in the private non-academic sector. Projections of demand for STEM graduates in the EU job market indicate a deficit of this kind of highly skilled labourers. The 2015 EU Commission report highlights that: “due to demographic developments, there will be a high replacement demand for high-skilled professionals working in STEM-related occupations in the coming years. This has led to concerns that Europe could lack an adequate supply of STEM skills to enable its future economic development”\(^2\).

**Non-academic career path**

STEM graduates, especially ones with advanced degrees (e.g. PhD), have a proven track record of being productive contributors in many sectors, from education to technology and finance. The DESCARTES consortium aims at equipping the ESRs with a combination of strong technical skills (e.g. statistics, coding, mathematics) and analytic-thinking skills particularly suitable to tackle complex problems often faced in modern business organisations.

Academic training curricula often lack sufficient exposure to techniques, technologies and business acumen necessary for attractive profiles in business and industry (PhD profiles are often labeled as being “too academic” by recruiters). In order to bridge this gap, we plan to exploit the synergy between DESCARTES academic and industrial partners in order to develop competitive professional profiles for the future job market. This is implemented through the following types of activities:

- Secondments with non-academic partners in order to exchange knowledge between academia and the private sector.
- Joint ESR supervision schemes for non-academic and academic beneficiaries.
- ESR training events (summer schools and workshops), to allow for efficient sharing of knowledge and experiences throughout the program.

Additionally, within their research programs, the ESRs will gain in-depth expertise about methodologies and technologies in high demand by the technology industry. These include, for example, data-science analysis, advanced machine-learning algorithms, modern software stacks for "big-data" analysis handling, firmware programming for FPGA hardware. The originality of the DESCARTES proposal consists in exposing the ESRs to a genuinely interdisciplinary environment allowing them to appreciate the potential impact of these techniques in other application fields.

All of the above are essential for improving the perspectives for successful employment of ESRs. We hence see the participation of both academic and non-academic partners essential for the success of this program.

**Academic career path**

DESCARTES ESRs will perform cutting edge research on topics within highly active fields of statistics and fundamental physics. Their work will attack questions with potentially significant impact on science (e.g. novel statistical inference methods, searches for new physics …), in order to maximise the quality of their research portfolios. ESRs will also operate and have visibility in large LHC scientific collaborations like those of the Atlas, CMS and LHCb experiments.

In addition to developing a strong academic portfolio, the combination of intense academic research with experiences and practices prevailing in business and industry will help produce more attractive candidates for academic positions. Many techniques and technologies ESRs will gain expertise with during their placement with non-academic partners can have a fruitful impact on research in HEP. Nevertheless, the adoption of these techniques by academia often takes a very long time: this is an area in which the DESCARTES program can have a significant impact also through the dissemination program within the CERN scientific collaborations. Co-mentoring and placement of students within both academic and non-academic environments aids in producing candidates with a much broader skill set and hence provide an edge when looking for a research career in HEP.

\(^2\) “Does the EU need more STEM graduates?”, EU Comission, 2015.
The HEP Software Foundation white paper \(^3 \) outlines the importance software development plays in constructing and maintaining core systems for detector triggers, HEP simulations, and data analysis, as well as challenges future of HEP will face due to a lack of physicists with advanced skills in software development suitable for such roles. DESCARTES, through synergy between academic and industry partners, offers an excellent opportunity to produce more profiles to support the future of HEP software development. DESCARTES already incorporates projects directly linked to the support of critical triggering and event tagging software, perfectly in line with the needs outlined in the HEP Software Foundation white paper.

Furthermore, as machine learning becomes more and more adopted in HEP experiments and data analysis, it is important that the demand for profiles of physicists proficient in machine and deep learning techniques can meet the demand. Secondments and placement with non-academic partners, DESCARTES training programs as well as the co-mentoring scheme we have envisioned will aid in producing such profiles.

**Skills**

A standard path towards a PhD degree in HEP implies development of many transferable skills. These include but are not limited to:

- Advanced statistics (hypothesis testing, monte-carlo techniques …)
- Advanced mathematics (linear algebra, advanced calculus, functional analysis, probability…).
- Coding (C++, Python …).
- Data Engineering (data wrangling, feature engineering, data simulation…).
- Analytic thinking skills.
- Strong verbal and written communication skills (scientific document writing, presentation design and delivery…).

Participation of ESRs in HEP projects will allow them to develop the above technical skills to an advanced degree, both through direct research and trainings provided by DESCARTES. Writing research papers, internal collaboration notes, and participation in conferences/workshops will contribute to developing strong communication skills.

Despite strong fundamentals in math, statistics, coding and research characteristic of HEP PhDs, entry into non-academic fields is often made difficult by the negative bias of business oriented organisations towards profiles which are deemed “too academic.” In order to bridge this gap, and make DESCARTES ESRs more attractive on the job market, it is essential that the above mentioned skills are also complemented by skills such as:

- Modern Project management (Agile, Waterfall…) and time management.
- Soft skills associated with working in non-academic environments (industrial domain knowledge, work in non-academic teams…)
- Awareness of legal frameworks such as intellectual property rights and GDPR.
- Business development skills (Client interaction, business acumen…).
- Modern coding paradigms/practices (asynchronous coding, versioning, proper documentation …).
- Big data techniques (Apache Spark, Hadoop, …).
- Modern software stacks (pandas, numpy, sklearn, tensorflow, pytorch…).

Participation in commercial projects, either through secondments or through full time placement, will allow ESRs to develop the skills necessary for:

- Successful potential careers in business and industry.
- Transferable knowledge and experience from the business sector useful for an academic research career.

As machine and deep learning are central themes in the DESCARTES project, it is worth emphasising that most recent developments in deep learning technologies originate from non-academic organisations (e.g. reinforcement learning from DeepMind). Such developments are quickly adopted by the non-academic sector, but a latency for adoption within academia for novel deep learning technologies exists. We hence believe that participation of industry partners in DESCARTES is of great potential for further development of students’ skills with regards to modern technologies which are common in business and industry but not as common in academia.

**2.2 Contribution to structuring doctoral/early-stage research training at the European level**

The network will build upon the proven and successful experience of the participating nodes in providing a traditional Doctoral-level training, to offer the ESR with an innovative program, with the long-term goal of having an impact on the European Ph.D. level education. The driving idea is to shape the training around the paradigm of interdisciplinarity, leveraging on the inherent characteristics of the data science field to bring together statisticians,

---

computer scientists, and domain scientists (HEP) for cross-domain fertilization. This goal is pursued through all the three pillars of the ESR preparation: the training through research, the participation in dedicated schools and training events, the participation in conferences and secondments:

- The research projects have been devised to foster the synergy between the statistical research topics and the HEP related ones.
- Specific training opportunities are meant to connect the ESR with other application fields using similar or complementary statistical tools or having to deal with similar use cases. This is implemented exploiting the existing connections of the institutes involved in the research in statistics with other application domains, and via the non-academic partnerships.
- The ESR will report and participate in conferences beyond the HEP domain to keep the pace with the quickly evolving expertise in the domain of statistical tools and machine learning.

A further opportunity to foster the creation of a truly diverse skill-set is implemented within the HEP related part of the research program, exposing the trainee to both fundamental and applied topics. This is true for the achievement of the ITN's deliverables as well as for the ESR's work in the context of the large CERN LHC experimental collaborations; the ESRs will have to participate in the operation of their experiment of affiliation to be granted authorship rights.

The network design aims at maximizing the impact of the non-academic nodes on all aspects of the ESR training. Both the non-academic beneficiaries and partners will have an active role in the training through research, also via the implementation of a balanced secondment plan. Moreover, they will actively participate in formal training initiatives contributing to the enrichment of the ESR experience with skills in high-demand by the job market. Concretely, to give some examples, companies like Achilles and Artabro will expose the ESR to the utilization of novel statistical-learning techniques for non-academic use-cases like Industry 4.0 and Business Intelligence services. The ESRs collaborating with B12 will work on the development and utilization of data-science frameworks (like the Spark analysis ecosystem) with potential applications also in the private sector. The expertise of partners like TWave in the realm of software development and FPGA firmware programming is highly complementary with the work of the ESRs on the development of real-time algorithms for medical and HEP applications.

Besides enriching the ESRs portfolio of assets, this collaboration strategy is also meant to increase the attention of the academic sector towards novel technologies already in use in the non-academic environment. In the long term, such an open attitude by researchers will have a positive and long-lasting impact on the training and competitiveness at the European level. The CERN ecosystem, with its international collaborations, also represents a unique opportunity for early-stage researchers to experience a truly diverse working environment. The ESRs will be called to collaborate and report back to the experiment their research progress expanding further their potential for training through research and relating to colleagues with different nationalities, expertise, and background.

This continuous feedback between the network members and the greater HEP community is an essential asset of the network to have an impact on the European level Doctoral education beyond the span of the ITN itself. The bonds between researchers in HEP and statistics, the acquired expertise with tools employed in other application areas, and the private sector will disseminate and percolate on the broader community. They will have a beneficial impact on the research output and employability of future Doctoral students. The training offer and dissemination strategy have been devised to maximize this virtuous impact, for example, advertising and opening a part of the research programs to participation outside the network.

2.3 Quality of the proposed measures to exploit and disseminate the results

2.3.1 Dissemination of the research results

The dissemination strategy of the network fulfills a twofold objective: on one side, it aims at maximizing the exploitability of the scientific results and, on the other, it endeavors at having an impact on the Doctoral level training beyond the participating nodes. Three main paths will be followed: academic dissemination to the broader scientific community, dissemination in the non-academic sector, dissemination through interaction within the CERN HEP collaborations and network partners. The latter being particularly relevant considering the size of the LHC experimental collaborations, which are composed of several thousand scientists, including several hundreds of Doctoral students. As for the former two, special attention will be devoted to widespread scientific results outside the HEP community, via the development of statistical and computation methods of benefit for other science domains. The specific activities are presented in Table 2.3.
Table 2.3 Dissemination strategy

<table>
<thead>
<tr>
<th>Academic dissemination to the broader scientific community</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Report at workshops and international conferences.</strong> These contributions will concern both the physics results obtained on the LHC data and the statistical and applicative topics. Particular emphasis will be given to the interdisciplinarity aspects of the research program via the development of statistical methodologies applicable to different domains where similar statistical issues arise.</td>
</tr>
<tr>
<td><strong>Publication of the HEP analysis results in high impact peer-reviewed scientific journals.</strong> The consortium will adhere to the Open Access publication policy of the CERN LHC experiments to make the results available to the broadest possible audience.</td>
</tr>
<tr>
<td><strong>Publication about new software tools, statistical algorithms and models, and other applied topics</strong> of the research program in specialized scientific journals.</td>
</tr>
<tr>
<td><strong>Distribution of software products</strong> as standalone or as part of more extended software libraries (as hls4ml or experiment software) depending on the use-case.</td>
</tr>
</tbody>
</table>

**Dissemination in the non-academic sector**

| Organization of a **thematic workshops open to non-academic and private companies outside the network** in collaboration with the CERN Knowledge-Transfer hub⁴ to present network results with potential for application in the private domain. |
| **Dissemination through interaction within the CERN HEP collaborations and network partners.** |
| Presentation of the research results and the methodology related aspects in LHC experimental **collaborations meetings** and workshops, and ITN network-wide events. Seminars and colloquia in the participating institutes. |
| **Publication of the research results and the methodology related aspects in internal documents to the LHC experimental collaborations** (Atlas and CMS notes). |
| Software implementation, distribution, and integration with the experiment analysis and event reconstruction frameworks for fruition by all members of the collaboration. |

2.3.2 Exploitation of results and intellectual property

The dissemination strategy of the ITN network towards the broader HEP community follows a consolidated path for this research field, thus ensuring the exploitation of the scientific output of the network. New public results rapidly contribute to the evolution of the HEP knowledge prompting new research or the creation of new models by theoreticians and phenomenologists. The publication of the detailed description and, when foreseen, of the software used for the methodological aspects, the applied physics ones, and the statistical algorithms will also allow them to be adopted by the scientific community. Having a dedicated dissemination strategy internal to the CERN experimental collaborations and establishing close collaboration with other experiment members will guarantee an even faster adoption of the ITN results and research methods within these scientific communities. Methodologies and software tools will be potentially adopted for studies and analysis beyond those proposed by the network.

Publication using the data of the CERN experiments, like those documenting the searches and analysis performed by the ITN researchers, will have to follow the authorship rules of the corresponding collaboration and be signed by their full author list. Public reports, documentation, and publications describing progress in specific application areas, such as advanced theoretical framework or new statistical learning methods, will be signed only by the authors of the work. The software will be released following open-source policies unless developed during ESR secondments with industrial partners, in which cases it will be subject to the intellectual property rights of these companies. FIXME: this needs to be addressed, what are the policies for the SW in the specific case of the participating companies? FIXME: all examples need to be reviewed once the text of the project is refined. FIXME: This has to focus on impact outside the academic world.

2.4 Quality of the proposed measures to communicate the activities to different target audiences

2.4.1 Communication and public engagement strategy

The physics reach of the LHC and its success as a truly international scientific collaboration brings a huge inspirational value. We will train our ESRs to be excellent science communicators and they will play a leading role in communicating the goals of DESCARTES to various audiences using different methods. They will:

1) write blogs highlighting their research and the excitement of fundamental research;
2) promote their research to the public at events such as the European Researchers’ Night (ERN) and La Fête de la Science;
3) participate in events such as Pints of Science and Café Scientifique which target audiences not usually engaged in science outreach activities;

⁴ https://kt.cern/
4) convey the importance of science to students of different age groups by giving at least one talk a year at a school and by helping with the International Master Classes at their institution;
5) develop computer applications using virtual reality to give the public and younger students a more direct idea of the complexity of the LHC detectors and events.

The network will also hold events focusing on public engagement during our yearly meetings. We will invite a world-leading researchers in our field to give a keynote public lecture that will be well publicised. We will also invite the audience to a small poster session where all ESRs will display posters describing their research to a non-specialist audience.

3. Quality and Efficiency of the Implementation

3.1 Coherence and effectiveness of the work plan

Table 3.1a provides detailed information on the work packages.

Table 3.1a: Description of Work Packages

<table>
<thead>
<tr>
<th>WP Number</th>
<th>WP Title</th>
<th>Months 1 – 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Searches for new physics in LHC data</td>
<td></td>
</tr>
</tbody>
</table>

**Lead Beneficiary**

UCA (J. Donini)

**Objectives**

- Develop a framework for global new physics searches based on efficient background modelling alternative to full simulations and semi-supervised learning algorithms for signal detection.
- Develop a framework for optimised searches for exotic long-lived particles in final states involving displaced leptons and reconstructed top quarks.
- Study the sensitivity to boosted di-Higgs boson production in HL-LHC in the context of new physics scenarios using advanced ML algorithms for particle identification in boosted jets.
- Search for new physics signal in b/anti-b quark forward-backward production asymmetry using AI techniques for jet flavour- and charge-tagging.

**Description of Work and Role of Specific Beneficiaries / Partner Organisations**

WP1 focuses on searches for new physics at the LHC using novel and advanced statistical inference tools. Two strategies are applied: model-independent searches for anomalies in the expected behaviour of large data sets, possibly pointing to a presence of a new physics signal, and searches for special signatures in final states that can unveil new physics. The work is divided into three tasks:

**Task 1: Global search for new physics at the LHC.** Two approaches are considered: (i) New physics searches usually rely on vast amounts of time-consuming and resource-intensive simulations. Alternative, multivariate density estimation methods of efficient and accurate background modelling in searches for new physics will be developed. Novel approaches will be explored, e.g. based on gaussian processes, nonparametric likelihood, or NN-based algorithms. (ii) A model-independent search for new physics will be performed in the context of anomaly detection via semi-supervised learning, either by training unsupervised methods (e.g. autoencoders, GANs, or triplet networks) or via the suitable adjustments of supervised approaches (e.g. BDT or NN) to specifically account for the rarity of the searched signal. UCA on the physics side and UNIPD on the statistics side will play the key roles in this task.

**Task 2: Data-driven model-independent search for long-lived particles at the LHC.** The existence of long-lived particles (LLP) is a direct consequence of many BSM extensions. An analysis strategy will be set to perform a data-driven search for LLP, and it will be optimised independently for Run II and Run III data and for the HL-LHC. Autoencoders (AE) will be trained directly with data on long-lived particles associated with SM processes, which leave characteristic signatures involving objects significantly displaced from the primary vertex of the event when they interact with the detector components. Such patterns can be read out and used as input for the AE, which then have a high chance to identify any LLP that deviates from these patterns as a potential signal event. To evaluate their sensitivity, LLP candidates from various well-motivated SM extensions will be considered and exclusion limits on large classes of models will be set. IASA and AI4B will lead this task.

**Task 3: Particle algorithms for new physics searches.** High transverse momenta topologies offer a good handle to explore boosted objects and investigate BSM scenarios. Advanced machine learning algorithms will be developed for improving jet substructure image resolution and tagging efficiency with the aim of (i) investigating contact interactions predicted in BSM scenarios that enhance di-Higgs boson (HH) production cross section at high H1 mass values, (ii) improve jet flavour and charge tagging for the measurement of the b/anti-b forward-backward production asymmetry where a deviation from the SM prediction is a clear indirect signature of new physics. This task will be driven by LIP and INFN for the physics studies and by UNIPD and UiO for the statistical algorithm developments. Synergies are envisioned with WP2, where the algorithms can be directly tested on data, and with WP4 where they may be implemented in the event reconstruction and classification at the HL-LHC.

---

7 For a review, see www.deeplearningbook.org, chapter 14.
8 https://arxiv.org/abs/1406.2661
Objectives

- Study the sensitivity to the Yukawa coupling of the H decay to cc-bar in Run II and III and in HL-LHC by using deep learning algorithms for jet substructure identification and software real time selections based on ad-hoc algorithms for efficient H(cc) candidate selection.

- Improve the reconstruction efficiency of H(bb) and H(ττ) decays and study the sensitivity to the Higgs boson self-coupling in LHC Run II, Run III and in HL-LHC data taking by using deep learning algorithms for t- and b-jet identification, event reconstruction and classification.

- Establish the discovery of ttH production and reach sensitivity to find evidence for tHq production in multi-lepton final states using advanced ML algorithms for background modelling and event classification.

- Develop advanced machine learning algorithms for shower reconstruction and classification from data with 3D spatial resolution and for Higgs boson measurements with very low S/B fraction.

Description of Deliverables. Numbers in parentheses indicate the month of delivery.

D1.1 Document describing the study of LLP signatures and deployment of the search strategy for Task 2 (18).
D1.2 Software and documentation for the statistical algorithms used in background modelling in Task 1 (24).
D1.3 Document describing the development and validation of the search method for Task 1 (24).
D1.4 Document describing the deployment of learning algorithms in the search strategy of Task 2 (30).
D1.5 Document describing the study of boosted signatures using deep learning algorithms (30).
D1.6 Document describing the b/anti-b asymmetry study using ML algorithms for jet charge tagging (42).

WP Number

<table>
<thead>
<tr>
<th>WP Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurements of Higgs boson properties with LHC data</td>
</tr>
</tbody>
</table>

Lead Beneficiary

INFN (D. Lucchesi)

WP Title

Effective field theories for physics at the LHC

Lead Beneficiary

UOXF (D. Bortoletto)

Objectives

- Develop a framework with the potential to provide a complete scan of the EFT parameters in the Higgs sector using deep learning tools.

● Constrain EFT parameters with Higgs boson measurements using simulation-based inference.
● Develop a framework for analysis of systematic uncertainties and improved interpretability of deep learning models applied in HEP and beyond.
● Provide a complete analysis paradigm for Higgs boson studies with distributed computing based on Apache Spark.

Description of Work and Role of Specific Beneficiaries / Partner Organisations

WP3 aims at probing new physics in a way that is complementary to the direct searches of WP1, by constraining Effective Field Theory (EFT) parameters from LHC data. The proposed approach supersedes previous matrix element-based methods and makes use of deep learning algorithms. It systematically examines inferential and other sources of uncertainties and it explores modern tools of efficient computing required for CPU-intensive analyses. The work plan involves three tasks:

Task 1: Deep learning algorithms for EFT parameter constraints from experimental data. The recent augmented data method\textsuperscript{16}, yet in the proof-of-concept stage, will be further developed to constrain EFT parameters in the Higgs boson sector, extending interpretations by the LHC collaborations in the EFT context\textsuperscript{17} which are restricted to the study of a predefined set of benchmark scenarios. The performance of the new method will be compared with more direct approaches based on the matrix element evaluation\textsuperscript{18}, either directly or via an ansatz in the form of a parametric DNN. Differential Higgs boson measurements will be optimised for the extraction of direct and indirect constraints on the Higgs boson self-coupling. UCL and UOXF will pursue this task, with a contribution from UMIB in the development of the augmented data method.

Task 2: Advanced inference tools for EFT studies. Novel pseudo-likelihood functions will be developed to approximate intractable likelihood functions and form directly a composite likelihood ratio for constraining EFT parameters from experimental data. Composite likelihood ratios provide inferential procedures relatively insensitive to systematic uncertainties. The effectiveness of the novel pseudo-likelihoods will be essayed for hypothesis testing in cognitive neuroscience experiments as well. The predictive accuracy of learning algorithms in the context of EFT will be studied and new methods will be developed based on conformal inference\textsuperscript{19} to construct provably valid prediction intervals for these algorithms. This task will be driven by UMIB in collaboration with UOXF and UCL.

Task 3: Computing and inference utilities for LHC analyses. Two sub-tasks are considered: (i) An advance in computing technology for HEP, where a modern distributed computing framework based on Apache Spark\textsuperscript{20} will be implemented for the purpose of streamlining LHC data analyses and improving the overall speed and efficiency of numerical computations. A case study of EFT in the Higgs sector will be performed utilising the Spark-based CERN infrastructure (SWAN, Hadoop-YARN, Kybernetes) to demonstrate the utility of the new framework and compare it with the existing distributed computing frameworks. (ii) An important development in inference methodology, where pertinent techniques such as bootstrapping will be explored to study systematic and inferential uncertainties of deep learning algorithms in the EFT context. Feature importance methods such as SHAP\textsuperscript{21} will be explored to understand the predictive power of the algorithms and improve their interpretability. B12, UMIB, and OSLO will pursue this task.

Description of Deliverables. Numbers in parentheses indicate the month of delivery.

D3.1 Software implementing Apache Spark for improved efficiency of distributed computing in Task 3 (18).
D3.2 Document reporting the study of H(bb) production using ATLAS data for Task 1 (42).
D3.3 Document reporting the study of H production using CMS data for task 1 (42).
D3.4 Document describing the framework for EFT studies in Task 1 (36).
D3.5 Document describing the framework for EFT studies in the Higgs sector in Task 1 (36).
D3.6 Software implementing and document describing the framework for systematic uncertainty studies of deep learning models in Task 3 (36).
D3.7 Document reporting the study of H(bb) production using ATLAS data for Task 1 (42).
D3.8 Document reporting the study of HH production using CMS data for task 1 (42).

WP Title
Real-time algorithms for data selection and medical imaging

Lead Beneficiary
CERN (G. Cerminara)

Objectives
● Develop deep learning models for particle reconstruction, identification, and selection for HL-LHC L1 triggers.
● Develop deep learning algorithms for medical image reconstruction in γ- and e⁻-emission tomography.
● Deploy L1 trigger and medical imaging algorithms on FPGA boards and demonstrate their performance.

Description of Work and Role of Specific Beneficiaries / Partner Organisations

The prospects for the HL-LHC program of the new physics searches and Higgs studies proposed in WP1 and WP2 strongly depend on the trigger capabilities of the upgraded detectors. WP4 focuses on developing advanced reconstruction and selection algorithms for deployment on custom trigger electronics (Task 1). This task requires leveraging recent advancements in libraries for the synthesis of machine-interpretable firmware starting from high-level programming languages to develop a reusable set of tools for the development, optimization, and deployment of complex algorithms on FPGA chips. These tools and methodologies will also be applied to reconstruction and classification techniques in medical tomography (Task 2).

Task 1: Fast algorithms for the Level-1 trigger upgrade for HL-LHC. The aim is to exploit the unprecedented high granularity of the upgraded CMS Level-1 trigger read-out and the flexibility of its custom electronics to improve the energy resolution and reconstruction efficiency of the online algorithms. In particular, the CMS high-granularity calorimeter will allow reconstructing image-like snapshots of the energy deposit\textsuperscript{22}. The additional challenge for the L1 trigger is running the inference on programmable hardware and with very short

Deep learning algorithms for particle shower reconstruction (inference of shower energy), identification (classification of impinging particle type), and selection will be developed and optimised for fast inference on FPGA chips. Their performance will be evaluated using simulation, hardware emulators, and demonstration boards. CERN and KROMEK will pursue this task.

Task 2: Fast ML algorithms for medical image reconstruction. Deep learning algorithms have rapidly become a methodology of choice for reconstructing medical images. Many modalities in tomography could potentially improve their performance by introducing image reconstruction techniques based on AI. Unsupervised deep learning algorithms for tomographic image reconstruction from emission projection data will be developed. Reconstruction results with experimental phantoms and clinical data will be compared and tested for both SPECT (Single Photon Emission Computed Tomography applied in both γ-Camera and Compton-Camera devices) and PET (Positron Emission Tomography) modalities in collaboration with hospitals and other health diagnostic institutions. The new algorithms will be adapted to firmware implementation and their performance will be evaluated on programmable hardware emulators. IASA and KROMEK will pursue this task.

Task 3: Algorithms for information reconstruction from radiation detectors in medical imaging. When a particle interaction is recorded in a detector, the resultant signal is a result of the detector transform acting upon the input signal. This transformation can be complex, dependent upon a number of factors, such as input energy, positional information, charge effects and contribution from the readout electronics. Machine learning offers the potential to recover some of the input information from the output signal, leading to improve energy and spatial resolutions, and improve materials discriminations and signal level. A mixture of experimental and simulated data will be created, and used to determine the optimum strategies to apply ML to this task, and the subsequent image reconstruction. Kromek will pursue this task.

Description of Deliverables. Numbers in parentheses indicate the month of delivery.

D4.1 Document describing selected algorithms and the metric for their performance evaluation on simulated data for Tasks 1 and 2 (18).
D4.2 Software tools and document reporting the prototype implementation and requirements for the firmware deployment of Tasks 1 and 2 (30).
D4.3 Document describing the algorithm performance and the deployment strategy in the upgrade architecture for Task 1 (42).
D4.4 Document describing the fully deployed prototype for background rejection and optimal resolution in Task 2 (42).
D4.5 Document describing strategies for application of ML to detector signal reconstruction in Task 3 (30).
D4.6 Document describing the algorithm performance and the deployment strategy for medical imaging for Task 3 (42).

WP Number 5  Months 1 – 48
WP Title Training and career development
Lead Beneficiary UCL (C. Delaere)

Objectives
- Identify analogies between problems in HEP and in other applications of advanced statistical inference, to foster synergies between HEP and other fields.
- Develop the ESRs’ ability to export advanced statistical methods from HEP to other domains.
- Provide experience of different methodologies and environments in participating companies.
- Award doctoral degrees to ESRs engaged in doctoral programs.
- Develop tailored career paths for ESRs to enhance their talent and direct them to successful career choices.

Description of Work and Role of Specific Beneficiaries / Partner Organisations
WP5 addresses the issue of creating customised training paths for the ESRs involved in the network, as well as finding non-academic applications of the developed products. The design of an individual career development plan of the trainees will involve all supervisors and it will be based on the detailed development of a Personal Career Development Plan (PCDP) for each ESR. This will specify the most appropriate training courses for each trainee, define the research objectives, finalise the secondment offer better suited and in best synergy to the ESR research plan, and design the skeleton of their doctoral thesis.

A different goal is to find appropriate use cases enabling the ESRs to propose applications to non-HEP problems of the techniques developed in their research work. A tight connection with partner organisations will be established through seminars at network-wide events.

This WP involves all ESRs and their academic and non-academic supervisors and co-supervisors. ESRs will be helped and encouraged by network members to find applications of the products developed within the network to the problems they work on during their secondments at partner organisations. The know-how and expertise they will gain in the secondments will be shared with the network members through seminars at network-wide events.

The activities of WP5 are divided in two main tasks. The first is the training of the ESRs. They will receive training and will acquire skills and know-how that can be used in the job market. This will occur within the beneficiary organizations through supervision, courses, seminars, interaction with visiting scientists; and at workshops, schools, and in secondments at partner institutions. The second is the definition of synergies in secondments at partner organisations: the secondments will also allow trainees to identify opportunities for technology transfer from and to the academic sector. All partners have an important role in these tasks, by providing non-HEP use cases and commercial applications to be developed during the secondments.

Description of Deliverables. Numbers in parentheses indicate the month of delivery.

D5.1 Design of Personal Career Development Plan (PCDP) for each ESR (8-16, depending on ESR hiring).
D5.2 Training in soft skills (20).
D5.3 Network seminars to share the expertise gained in secondments in industrial partners (40).

WP Number 6  Months 1 – 48
WP Title Dissemination and outreach activities

Table 3.1b: Deliverables list

<table>
<thead>
<tr>
<th>Scientific Deliverables</th>
<th>Deliverable Title</th>
<th>WP No.</th>
<th>Lead Beneficiary Short Name</th>
<th>Type</th>
<th>Dissemination Level</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.1</td>
<td>Baseline method for new physics searches with displaced objects in CMS Phase-1 and Phase-2</td>
<td>1</td>
<td>IASA</td>
<td>report</td>
<td>private</td>
<td>18</td>
</tr>
<tr>
<td>D1.2</td>
<td>Multivariate method for background modelling in new physics searches at the LHC</td>
<td>1</td>
<td>UNIPD</td>
<td>software + report</td>
<td>public</td>
<td>24</td>
</tr>
<tr>
<td>D1.3</td>
<td>Unsupervised and semi-supervised method for new physics search</td>
<td>1</td>
<td>UCA</td>
<td>report</td>
<td>public</td>
<td>24</td>
</tr>
<tr>
<td>Deliverable Number</td>
<td>Deliverable Title</td>
<td>WP No.</td>
<td>Lead Beneficiary Short Name</td>
<td>Type</td>
<td>Dissemination Level</td>
<td>Due Date</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>--------</td>
<td>-----------------------------</td>
<td>------</td>
<td>--------------------</td>
<td>----------</td>
</tr>
<tr>
<td>D1.4</td>
<td>ML algorithms for long-lived particle searches</td>
<td>1</td>
<td>IASA</td>
<td>report</td>
<td>public</td>
<td>30</td>
</tr>
<tr>
<td>D1.5</td>
<td>Studies of boosted objects using ML tools</td>
<td>1</td>
<td>LIP</td>
<td>report</td>
<td>private</td>
<td>30</td>
</tr>
<tr>
<td>D1.6</td>
<td>Sensitivity studies of the b - anti-b production asymmetry with jet charge tagging at LHCb</td>
<td>1</td>
<td>INFN</td>
<td>report</td>
<td>public</td>
<td>42</td>
</tr>
<tr>
<td>D2.1</td>
<td>Sensitivity projections of ML algorithms for ttH and tHq analyses using the Run 2 and Run 3 datasets</td>
<td>2</td>
<td>OVD</td>
<td>report</td>
<td>private</td>
<td>24</td>
</tr>
<tr>
<td>D2.2</td>
<td>Boosting algorithms for imbalanced data</td>
<td>2</td>
<td>UiO</td>
<td>report</td>
<td>public</td>
<td>24</td>
</tr>
<tr>
<td>D2.3</td>
<td>Studies of leptonic Higgs boson decays using ML tools</td>
<td>2</td>
<td>LIP</td>
<td>report</td>
<td>private</td>
<td>30</td>
</tr>
<tr>
<td>D2.4</td>
<td>Unsupervised method for shower reconstruction</td>
<td>2</td>
<td>UNIPD</td>
<td>report</td>
<td>public</td>
<td>36</td>
</tr>
<tr>
<td>D2.5</td>
<td>Studies of Higgs boson event (trigger and objects) reconstruction</td>
<td>2</td>
<td>CERN</td>
<td>report</td>
<td>private</td>
<td>30</td>
</tr>
<tr>
<td>D2.6</td>
<td>Study of LHCb sensitivity on H-&gt;cc search using Run III data</td>
<td>2</td>
<td>INFN</td>
<td>report</td>
<td>public</td>
<td>30</td>
</tr>
<tr>
<td>D2.7</td>
<td>Combined measurement of ttH and tHq production using ML algorithms.</td>
<td>2</td>
<td>OVD</td>
<td>paper</td>
<td>public</td>
<td>40</td>
</tr>
<tr>
<td>D3.1</td>
<td>Apache Spark platform for efficient distributed computing in HEP analyses</td>
<td>3</td>
<td>B12</td>
<td>software</td>
<td>public</td>
<td>20</td>
</tr>
<tr>
<td>D3.2</td>
<td>EFT studies through likelihood scans in the Higgs sector using the matrix element method</td>
<td>3</td>
<td>UCL</td>
<td>report</td>
<td>public</td>
<td>18</td>
</tr>
<tr>
<td>D3.3</td>
<td>New pseudolikelihood functions to compute the likelihood ratio for EFT parameters</td>
<td>3</td>
<td>UNIMIB</td>
<td>software + report</td>
<td>public</td>
<td>30</td>
</tr>
<tr>
<td>D3.4</td>
<td>Augmented data techniques and pseudo-likelihood functions for EFT studies in the Higgs sector</td>
<td>3</td>
<td>UCL</td>
<td>report</td>
<td>private</td>
<td>36</td>
</tr>
<tr>
<td>D3.5</td>
<td>Framework for EFT studies in the Higgs sector</td>
<td>3</td>
<td>UOXF</td>
<td>report</td>
<td>private</td>
<td>36</td>
</tr>
<tr>
<td>D3.6</td>
<td>Framework for studies of systematics in HEP deep learning models</td>
<td>2</td>
<td>B12</td>
<td>software + report</td>
<td>public</td>
<td>36</td>
</tr>
<tr>
<td>D3.7</td>
<td>Hbb study with ATLAS data</td>
<td>3</td>
<td>UOXF</td>
<td>report</td>
<td>public</td>
<td>42</td>
</tr>
<tr>
<td>D4.1</td>
<td>Online algorithms for CMS L1 trigger &amp; medical imaging: description and performance on simulation</td>
<td>4</td>
<td>IASF</td>
<td>report</td>
<td>private</td>
<td>18</td>
</tr>
<tr>
<td>D4.2</td>
<td>Algorithms for reconstructing input information from detected radiation signals</td>
<td>4</td>
<td>KROMEK</td>
<td>software</td>
<td>private</td>
<td>30</td>
</tr>
<tr>
<td>D4.3</td>
<td>Online algorithms for CMS L1 trigger &amp; medical imaging: prototype firmware implementation, tools for deployment, and resource estimates</td>
<td>4</td>
<td>CERN</td>
<td>software + report</td>
<td>public</td>
<td>30</td>
</tr>
<tr>
<td>D4.4</td>
<td>Final report on online algorithms for PET and SPECT imaging</td>
<td>4</td>
<td>IASA</td>
<td>report</td>
<td>public</td>
<td>42</td>
</tr>
<tr>
<td>D4.5</td>
<td>Final report on ML algorithms for information reconstruction for medical imaging detectors</td>
<td>4</td>
<td>KROMEK</td>
<td>report</td>
<td>public</td>
<td>42</td>
</tr>
<tr>
<td>D4.6</td>
<td>Final report on online algorithms for CMS L1 Trigger upgrade</td>
<td>4</td>
<td>CERN</td>
<td>report</td>
<td>public</td>
<td>42</td>
</tr>
</tbody>
</table>

**Management, Training, Recruitment and Dissemination Deliverables**

<table>
<thead>
<tr>
<th>Deliverable Number</th>
<th>Deliverable Title</th>
<th>WP No.</th>
<th>Lead Beneficiary Short Name</th>
<th>Type</th>
<th>Dissemination Level</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5.1</td>
<td>Design of personal career development plans</td>
<td>5</td>
<td>UCL</td>
<td>other</td>
<td>private</td>
<td>8-16</td>
</tr>
<tr>
<td>D5.2</td>
<td>Soft skills training</td>
<td>5</td>
<td>UCL</td>
<td>PDE</td>
<td>private</td>
<td>20</td>
</tr>
<tr>
<td>D5.3</td>
<td>Network seminars to share expertise gained in secondments in industrial partners</td>
<td>5</td>
<td>UCL</td>
<td>other</td>
<td>public</td>
<td>40</td>
</tr>
<tr>
<td>D6.1</td>
<td>Network blog</td>
<td>6</td>
<td>UNIPD</td>
<td>PDE</td>
<td>public</td>
<td>4</td>
</tr>
<tr>
<td>D6.2</td>
<td>Network hashtag on Twitter</td>
<td>6</td>
<td>UNIPD</td>
<td>PDE</td>
<td>public</td>
<td>4</td>
</tr>
<tr>
<td>D6.3</td>
<td>Public seminars by ESRs in high schools and public events</td>
<td>6</td>
<td>UNIPD</td>
<td>PDE</td>
<td>public</td>
<td>24-44</td>
</tr>
<tr>
<td>D6.4</td>
<td>Workshop for assessment of the impact of outreach activities of the network</td>
<td>6</td>
<td>UNIPD</td>
<td>PDE</td>
<td>public</td>
<td>40</td>
</tr>
<tr>
<td>D7.1</td>
<td>Advertisement of ESR positions</td>
<td>7</td>
<td>IASA</td>
<td>ADM</td>
<td>public</td>
<td>2-12</td>
</tr>
<tr>
<td>D7.2</td>
<td>Data management plan</td>
<td>7</td>
<td>IASA</td>
<td>ADM</td>
<td>public</td>
<td>4</td>
</tr>
<tr>
<td>D7.3</td>
<td>Recruitment completion</td>
<td>7</td>
<td>IASA</td>
<td>ADM</td>
<td>public</td>
<td>12</td>
</tr>
<tr>
<td>D7.4</td>
<td>Organisation of network-wide events</td>
<td>7</td>
<td>IASA</td>
<td>ADM</td>
<td>public</td>
<td>10-40</td>
</tr>
</tbody>
</table>
The major milestones of the project are listed in Table 3.1c.

**Table 3.1c: Milestones list**

<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
<th>Related WP(s)</th>
<th>Lead Beneficiary</th>
<th>Due Month</th>
<th>Means of Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Framework for global new physics searches with prompt signatures available</td>
<td>1</td>
<td>UCA</td>
<td>18</td>
<td>D1.3</td>
</tr>
<tr>
<td>M2</td>
<td>Background studies and analysis strategy for long-lived particle searches</td>
<td>1</td>
<td>IASA</td>
<td>18</td>
<td>D1.1</td>
</tr>
<tr>
<td>M3</td>
<td>Framework for long-lived particle searches available</td>
<td>1</td>
<td>IASA</td>
<td>24</td>
<td>D1.4</td>
</tr>
<tr>
<td>M4</td>
<td>Choice and optimization of algorithms for boosted object identification</td>
<td>1,2</td>
<td>LIP</td>
<td>24</td>
<td>D1.5</td>
</tr>
<tr>
<td>M5</td>
<td>Choice and optimization of the algorithm for jet flavor charge determination</td>
<td>1</td>
<td>INFN</td>
<td>36</td>
<td>D1.6</td>
</tr>
<tr>
<td>M6</td>
<td>Apache Spark cluster</td>
<td>1,2,3</td>
<td>B12</td>
<td>18</td>
<td>D3.1</td>
</tr>
<tr>
<td>M7</td>
<td>Optimized framework for tH and tHq studies available</td>
<td>2</td>
<td>OVD</td>
<td>18</td>
<td>D2.1</td>
</tr>
<tr>
<td>M8</td>
<td>Optimized framework for Hττ studies available</td>
<td>2</td>
<td>LIP</td>
<td>18</td>
<td>D2.3</td>
</tr>
<tr>
<td>M9</td>
<td>Optimized reconstruction and selection algorithms for H-&gt;cc identification</td>
<td>2</td>
<td>INFN</td>
<td>18</td>
<td>D2.7</td>
</tr>
<tr>
<td>M10</td>
<td>Framework for jet identification based on quantum information algorithms</td>
<td>2</td>
<td>INFN</td>
<td>36</td>
<td>D2.6</td>
</tr>
<tr>
<td>M11</td>
<td>Reference method based on MEM available</td>
<td>3,2</td>
<td>UCL</td>
<td>18</td>
<td>D3.2</td>
</tr>
<tr>
<td>M12</td>
<td>Optimized framework for Hbb studies available</td>
<td>3,2</td>
<td>UOXF</td>
<td>18</td>
<td>D3.5</td>
</tr>
<tr>
<td>M13</td>
<td>Methodology based on pseudo-likelihoods for EFT parameters</td>
<td>3</td>
<td>UMIB</td>
<td>24</td>
<td>D3.3</td>
</tr>
<tr>
<td>M14</td>
<td>Augmented data technique available</td>
<td>3</td>
<td>UCL</td>
<td>36</td>
<td>D3.4</td>
</tr>
<tr>
<td>M15</td>
<td>Framework for systematic uncertainty studies in HEP deep learning models</td>
<td>1,2,3,4</td>
<td>B12</td>
<td>36</td>
<td>D3.6</td>
</tr>
<tr>
<td>M16</td>
<td>Clinical data and simulations for medical imaging available</td>
<td>4</td>
<td>IASA</td>
<td>18</td>
<td>D4.1</td>
</tr>
<tr>
<td>M17</td>
<td>Choice and optimization of algorithms for online CMS L1 trigger and medical imaging</td>
<td>4</td>
<td>CERN</td>
<td>18</td>
<td>D4.1</td>
</tr>
<tr>
<td>M18</td>
<td>Demonstration of algorithms for medical imaging and L1 trigger on FPGA boards</td>
<td>4</td>
<td>CERN</td>
<td>30</td>
<td>D4.3</td>
</tr>
<tr>
<td>M19</td>
<td>Demonstration of algorithms for input information reconstruction</td>
<td>4</td>
<td>KROMEK</td>
<td>30</td>
<td>D4.2</td>
</tr>
<tr>
<td>M20</td>
<td>Successful completion of statistics, ML and big data at UNIMIB, Uio and UCA</td>
<td>5</td>
<td>UCL</td>
<td>22</td>
<td>Web sites of the schools</td>
</tr>
<tr>
<td>M21</td>
<td>Successful completion of thematic workshops bridging industry and academia at CERN and KROMEK</td>
<td>5</td>
<td>UCL</td>
<td>39</td>
<td>Web sites of the workshops</td>
</tr>
<tr>
<td>M22</td>
<td>Successful completion of outreach workshops at LIP, CERN and OVD</td>
<td>6</td>
<td>UNIPD</td>
<td>47</td>
<td>Web sites of the workshops</td>
</tr>
<tr>
<td>M23</td>
<td>Award of doctoral degrees</td>
<td>6</td>
<td>UNIPD</td>
<td>48</td>
<td>Network blog</td>
</tr>
<tr>
<td>M24</td>
<td>Recruitment completion</td>
<td>7</td>
<td>IASA</td>
<td>12</td>
<td>D7.3</td>
</tr>
</tbody>
</table>

Table 3.1d describes the project title, the objectives and the expected results of the activity of each ESR, together with the planned secondment periods and their purpose. Variation in the secondments can happen due to progress of the activities of both the ESR and the hosting institution.

**Table 3.1d: Individual research projects**

<table>
<thead>
<tr>
<th>Fellow</th>
<th>Host institution</th>
<th>PhD enrolment</th>
<th>Start date</th>
<th>Duration</th>
<th>Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESR1</td>
<td>IASA</td>
<td>Y</td>
<td>6</td>
<td>36</td>
<td>D1.1, D1.4</td>
</tr>
</tbody>
</table>

**Project Title and Work Package(s) to which it is related:** Searches for exotic long-lived particles with the CMS detector using advanced learning algorithms (WP1, WP4).

**Objectives:**
1. Implement a preliminary cut-based analysis for searches for long-lived particles decaying into leptons.
2. Develop multivariate analysis with advanced learning algorithms to enhance the search sensitivity.
3. Optimize multivariate analysis for HL-LHC.

**Expected Results:**
1. Scientific publication of exclusion limits from Run-2 and Run-3 data, using the multivariate analysis developed for the search.
2. Technical note describing the analysis optimization and expected limits for HL-LHC.

**Planned secondment(s):**
1. UCA, supervisor J. Donini, acquire training on new physics search methods (months 13-14);
2. UNIPD, supervisor G. Menardi, work on D1.4 (months 25-26);
3. A14B, supervisor M. Spannowsky, acquire experience of industrial environment...
## Expected Results

- States at the CMS detector.

## Objectives

- Project Title and Work Package(s) to which it is related: Global searches for new physics with the ATLAS detector using semi-supervised algorithms (WP1).

## Planned secondment(s):

- (1) Padova, supervisor G. Menardi, work on D1.1, D1.2 (months 15-16); (2) UCI, supervisor D. Whiteson, acquire training on new physics search methods (25-26); (3) TWave, supervisor J. Blanco, acquire experience of industrial environment (months 37-39).

## Enrolment in Doctoral degree(s): NKUA; local supervisor: K. Vellidis.

<table>
<thead>
<tr>
<th>ESR2</th>
<th>UCA</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D1.3</th>
</tr>
</thead>
</table>

## Project Title and Work Package(s) to which it is related: Unexplored statistical approaches for the analysis of LHC data (WP1,WP2).

## Objectives:

- (1) Develop unsupervised methods for modeling backgrounds in searches for new physics. (2) Develop semi-supervised methods for anomaly detection in ATLAS data.

## Expected Results:

- (1) Scientific publication of semi-supervised methods for anomaly detection. (2) Technical note describing the non-parametric background modeling methods.

## Planned secondment(s):

- (1) Padova, supervisor G. Menardi, work on D1.1, D1.2 (months 15-16); (2) UCI, supervisor D. Whiteson, acquire training on new physics search methods (25-26); (3) TWave, supervisor J. Blanco, acquire experience of industrial environment (months 37-39).

## Enrolment in Doctoral degree(s): UCA; local supervisor: J. Donini.

<table>
<thead>
<tr>
<th>ESR3</th>
<th>UNIPD</th>
<th>Y</th>
<th>2</th>
<th>36</th>
<th>D1.2, D2.4</th>
</tr>
</thead>
</table>

## Project Title and Work Package(s) to which it is related: Jet identification with advanced Machine Learning algorithms at LHCb (WP1,WP2).

## Objectives:

- (1) Develop novel techniques for jet identification at the LHCb experiment exploiting jet substructure. (2) Develop high-level trigger for H->bb/cc selection and study the LHCb sensitivity on H->cc using Run 3 data. (3) Implement b vs anti-b classification to improve the b-cc asymmetry measurement.

## Expected Results:

- (1) Framework for jet substructure studies at LHCb. (2) Public report on novel machine learning approaches including quantum information algorithms on jet identification. (3) Scientific publication on the H->cc search using LHCb Run 3 data.

## Planned secondment(s):

- (1) UCA, supervisor J. Donini, work on D1.2 (months 17-18); (2) LIP, supervisor M. Gallinaro, work on D2.4 (months 23-24); (3) Achilles, supervisor F. Margaroli, acquire experience of industrial environment (months 31-33).

## Enrolment in Doctoral degree(s): UNIPD; local supervisor: G. Menardi.

<table>
<thead>
<tr>
<th>ESR4</th>
<th>INFN</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D1.6, D2.6</th>
</tr>
</thead>
</table>

## Project Title and Work Package(s) to which it is related: Machine Learning methods for new physics searches and HH->bbrr studies (WP1,WP2).

## Objectives:

- (1) Develop event classification methods for signal selection in Higgs boson studies. (2) Improve regression techniques for better Higgs boson mass resolution in events with τ leptons in the final state. (3) Study resonant di-Higgs production and constrain BSM models predicting resonant production.

## Expected Results:

- (1) Scientific publication on double Higgs production. (2) Technical note on boosted object identification.

## Planned secondment(s):

- (1) UiO, supervisor R. De Bin, acquire training on ML algorithms (months 13-14); (2) OVD, supervisor S. Folgueras, work on D2.3 (months 25-26); (3) B12, supervisor M. Backovic, acquire experience of industrial environment (months 37-39).

## Enrolment in Doctoral degree(s): LIS/IST; local supervisor: M. Gallinaro.

<table>
<thead>
<tr>
<th>ESR5</th>
<th>LIP</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D1.5, D2.3</th>
</tr>
</thead>
</table>

## Project Title and Work Package(s) to which it is related: Measurements of Higgs boson properties with LHC data (WP2).

## Objectives:

- Observation of tH production and first evidence of tHq production, where q is a light quark, in events with multi-lepton final states at the CMS detector.

## Expected Results:

- (1) Scientific publication on tH production. (2) Scientific publication on tHq production.

## Planned secondment(s):

- (1) IASA, supervisor K. Vellidis, work on lepton reconstruction and classification algorithms (months 13-14); (2) UiO, supervisor R. De Bin, acquire training on ML tools for rare signal detection (months 23-24); (3) B12, supervisor M. Backovic,
Objectives: (1) Develop gradient boosting algorithms for rare signal detection in very low-S/B samples of HEP data. (2) Develop selective inference tools and resampling-based methods to estimate systematic uncertainties of statistical learning algorithms in HEP.

Expected Results: (1) Scientific publication on boosting algorithms for analyses of highly imbalanced data. (2) Technical note on inferential tools for boosting algorithms. (3) Scientific publication on resampling techniques in machine learning.

Planned secondment(s): (1) OVD, supervisor S. Folgueras, work on D2.2 (months 15-16); (2) B12, supervisor M. Backovic, acquire experience of industrial environment (months 24-26); (3) INFN, supervisor D. Lucchesi, work on D3.6 (months 31-32).

Enrolment in Doctoral degree(s): OUD; local co-supervisors: J. Cuevas and S. Folgueras.

Objectives: Project Title and Work Package(s) to which it is related: Advanced statistical algorithms for HEP (WP2,WP4).

Expected Results: Enrolment in Doctoral degree(s): D3.6 (months 25).

Planned secondment(s): (1) CERN, supervisor G. Cerminara, work on interpretability of ML models in HEP (months 34-36).

Enrolment in Doctoral degree(s): UOXF; local supervisor: D. Bortoletto.

Objectives: Project Title and Work Package(s) to which it is related: Probing new physics with EFT (WP3).

Expected Results: Enrolment in Doctoral degree(s): D3.5, D3.7.

Planned secondment(s): (1) UNIMIB, co-supervisors A. Solari and N. Lunardon, work on D3.5 (months 15-16); (2) UCL, supervisor C. Delaere, work on D3.2 (25-26); (3) KROMEK, supervisor B. Cantwell, acquire experience of industrial environment (months 34-36).

Enrolment in Doctoral degree(s): UCL; local supervisor: C. Delaere.

Objectives: Project Title and Work Package(s) to which it is related: New algorithms for EFT studies (WP3).

Expected Results: Enrolment in Doctoral degree(s): D3.3.

Planned secondment(s): (1) UOXF, supervisor D. Bortoletto, work on D3.5 (months 17-18); (2) UCL, supervisor C. Delaere, work on D3.3 (months 28-29); (3) B12, supervisor M. Backovic, acquire experience of industrial environment (months 30-32).

Enrolment in Doctoral degree(s): UNIMIB; local co-supervisors: A. Solari and N. Lunardon.

Objectives: Project Title and Work Package(s) to which it is related: High-performance deep learning for HEP physics analyses (WP1,WP2,WP3).

Expected Results: Enrolment in Doctoral degree(s): UCL; local supervisor: A. Giammanco.
3.2 Appropriateness of the management structures and procedures

3.2.1 Network organisation and management structure

Figure 4 shows the organisation and management structure of DESCARTES. The project is organised in three levels: the management, the network activities, and the research activities. The management is performed by two bodies: a Network Management Committee (NMC) and a Supervisory Board (SB). The NMC consists of a Network Coordinator (Konstantinos Vellidis), a Scientific Coordinator (Daniela Bortoletto), a Training Coordinator (Christophe Delaere) and a Knowledge Transfer Coordinator (Giovanna Menardi). The SB consists of all beneficiary supervisors, chaired by Donatella Lucchesi. The implementation of the network management will be consolidated with a Consortium Agreement before the start of the project.

The role of the NMC is to continuously coordinate the implementation and monitor the progress of the project. This body holds regular monthly video-conference meetings, or more frequently if the need arises, and may call for a special SB meeting at any time if a strategic decision is required. The Network Coordinator chairs the NMC and coordinates the actions of WP7 involving the overall management of the project. He is responsible for the Financial Management, he oversees IPR aspects of the project, and he is in regular contact and cooperation with the EC Project Officer. The Scientific Coordinator is in charge of the continuous monitoring of the research programme in collaboration with the WP leaders, the timely release of the scientific deliverables, the verification and evaluation of the research milestones, and the coordination among the scientific WPs. The Training Coordinator is in charge of the continuous monitoring of the ESR training and coordinates the actions of WP5.

<table>
<thead>
<tr>
<th>ESR12</th>
<th>CERN</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D2.5, D4.3, D4.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title and Work Package(s) to which it is related: Lepton reconstruction and selection algorithms for the HL-LHC CMS Level-1 trigger upgrade (WP2, WP4).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives: (1) Study the acceptance needs of critical analyses for CMS HL-LHC run with a particular focus on the Higgs sector. (2) Develop algorithms for lepton triggers in HL-LHC exploiting energy-flow techniques and the high-granularity read-out of the CMS upgraded Level-1 trigger. (3) Evaluate the performance and resource consumption of selected algorithms for implementation in firmware on the HL-LHC CMS trigger.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Results: (1) Technical note on the physics potential of the proposed trigger algorithms. (2) Scientific publication on the algorithm performance, their implementation in firmware, and their deployment on demonstrator electronic boards.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planned secondment(s): (1) IASA, supervisor S. Stiliaris, work on D4.2 (months 15-16); (2) FNAL, supervisor S. Jindariani, work on D4.4 (months 25-26); (3) KROMEK, supervisor B. Cantwell, acquire experience of industrial environment (months 37-39).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolment in Doctoral degree(s): OVD; local supervisor: B. Alvarez.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ESR13</th>
<th>IASA</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D4.1, D4.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title and Work Package(s) to which it is related: Machine learning methods in Medical Tomography for SPECT and PET image reconstruction (WP4).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives: (1) Development of machine learning techniques to solve the inverse Radon Transform for Emission Tomography with simulated data. (2) Verification of the image quality with projection data from IASA’s γ-Camera system and from clinical SPECT and PET devices. (3) Implementation of the algorithms on custom electronic boards.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Results: (1) Technical note describing the methodology followed to solve the forward and inverse tomographic problem with machine learning techniques. (2) Scientific publication of a comparative study with simulated and real emission data demonstrating the performance of the developed methods.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planned secondment(s): (1) CERN, supervisor G. Cerminara, work on D4.1 (months 13-14); (2) FNAL, supervisor S. Jindariani, work on D4.2 (months 23-24); (3) KROMEK, supervisor B. Cantwell, acquire experience of industrial environment (months 37-39).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolment in Doctoral degree(s): NKUA; local supervisor: E. Stiliaris.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ESR14</th>
<th>KROMEK</th>
<th>Y</th>
<th>6</th>
<th>36</th>
<th>D4.2, D4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title and Work Package(s) to which it is related: Methods for detector spectrum and medical imaging reconstruction (WP4).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives: (1) Develop techniques for reconstructing input information from detected spectra, initially in semiconductor X- and gamma-ray detectors. Information to be reconstructed is likely to include input spectrum and increased positional resolution. (2) Application of the above methods, along with other techniques, to improve reconstruction in medical imaging applications.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Results: (1) Technical note describing the techniques for information recaptured from detection data. (2) Scientific paper outlining advances in reconstruction potential.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planned secondment(s): (1) INFN, supervisor D. Lucchesi, acquire training on ML for HEP (months 17-18); (2) IASA, supervisor S. Stiliaris, work on D4.3 (months 25-26); (3) CERN, supervisor G. Cerminara, work on D4.4 (months 37-39).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolment in Doctoral degree(s): University of Durham; local supervisor: M. Spannowsky.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ensures that the advertisement of the ESR positions, interviews, and hiring give equal opportunities to all candidates without regard to race, religion, color, national origin, sex, age, or disability. He also ensures the absence of gender bias in the workplaces of the ESRs, including their secondments. He oversees the PhD enrolment of the ESRs, the successful completion of their academic courses, the preparation and conduct of the network’s schools and training workshops, and the ESR path towards their doctoral degrees. He is also in charge of consulting with the ESRs for the deployment of their Personal Career Development Plans (PCDPs). The Knowledge Transfer Coordinator continuously monitors the publicity of the project and coordinates the actions of WP6. She is in charge of the network’s Press Office, creates the network’s blog, she updates it regularly with all public activities of the network, and she oversees the preparation and conduct of the network’s knowledge transfer and outreach workshops and seminars.

Figure 4: The management structure of the project. The arrows illustrate the interactions between the various components.

3.2.2 Supervisory board

The SB includes a representative of each beneficiary and partner organisation and one ESR representative. This body makes the strategic decisions concerning the project by majority vote of its members. The presence of the private sector in the SB ensures adequate balance between scientific training and transferable-skills training of the ESRs. The SB is responsible to establish active and continuous communication and exchange of best practice among the consortium members and oversee the quality and quantity of supervision of the ESRs. It meets approximately every six months, in coincidence with some of the main network events, or more frequently if the NMC calls for a meeting. In these meetings, the NMC reports to the SB on the implementation and progress of the project. Prof. Donatella Lucchesi will serve as the SB chair.

3.2.3 Recruitment strategy

Our experience indicates that there will be a wide pool of high-caliber candidates for the DESCARTES ESR positions. We will recruit the best possible candidates and follow the guidelines in the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers.

The recruitment of the ESRs requires an advertising campaign which will include:

- Circulating posters, announcements and e-mails to top research institutes and collaborations worldwide.
- Websites such as the EU portal (EURAXESS), inspirehep, findaphd.com, and jobs.ac.uk.
- Advertisement in social media accounts, such as Facebook and Twitter, of all participant organisations.

Advertising: All advertising will highlight the excellence of the research work, the conditions, the outstanding possibilities to work with competent companies through secondments and the required profile of the candidate. Special care will be devoted to ensure that all advertisements include a strong message toward gender equality with the statement “We strongly encourage the application by female students and researchers”. The posts will all be advertised shortly after the project has been accepted, but before its official start date, to ensure there is ample time to advertise and fill all the posts.
Application process- All applications and supporting documentation will be required to be in English, as this will be the working language of the network. Candidates will be required to provide a CV and relevant annexes, together with the names of 3 referees. The selection process will be rigorous. It will be based on scientific potential, academic qualifications, language knowledge, mobility and work ethic.

Selection- The Selection Board will be as gender-balanced as possible and will include at least the supervisor for the position, the Network Coordinator, and the Training Coordinator, with members from at least two European countries, and additional members from the partner organisations providing the secondments of the position. The Selection Board will produce a shortlist of candidates for each ESR post based on the application. The shortlisted candidates will be interviewed by the Selection Board. The result of the ranking will be discussed by the SB that will ratify the rankings before the top ranked candidates are notified of the outcome. Candidate selection will satisfy all recruitment regulations and will adhere to the code of conduct for the recruitment of researchers, as outlined in the European Charter for Researchers.

3.2.4 Progress monitoring and evaluation of individual projects

WP5 describes in detail the steps that will be taken in DESCARTES to monitor and evaluate individual projects. Each ESR will develop a PCDP in collaboration with their main supervisor and the Training Coordinator (TC). The Scientific Coordinator (SC) will track the progress in every research WP and report to the Network Management Committee (NMC), who will assess the status of individual projects during the monthly video-conference meetings. The WP leaders will provide the SC with brief updates on the respective WP progress and the supervisors will provide the TC with brief updates on the ESRs work for these meetings. The ESRs will give more complete presentations and the supervisors will report on the ESR progress regarding the institutional procedures for the doctoral degree program during the main Network meetings foreseen for months 10, 16, 24, 33, and 40. During these meetings, the WP leaders will provide status reports on the WPs and the NMC will evaluate the percentage of completion of the milestones and deliverables. The SB will discuss the progress of the project and of the ESRs and decide if steering actions should be implemented in coordination with the ESR supervisors. If progress is insufficient or hindrances are identified, the consortium will propose mitigation actions, e.g. by identifying volunteers among the network members who will provide added support to the ESRs in their projects.

3.2.5 Risk management

The proponents of DESCARTES have the expertise to deal with all the challenges of the research projects. Nonetheless, we have developed an adaptive plan of action to mitigate the risks in cases where a milestone cannot be reached technically or within the time limits of the network. WP leaders will report periodically to the NMC the progress within their WPs. If a WP leader cannot resolve a problem, the NMC will implement actions to keep the project on schedule. WP leaders will also envision a contingency plan to ensure the delivery of network products in case of an underperforming ESR, considering the WP interdependence. The possible risks that we have identified for our programme are listed in Table 3.2a, with outlines of the action plan we envision to manage each one of them, according to rules that will be detailed in the Consortium Agreement.

### Table 3.2a: Implementation risks

<table>
<thead>
<tr>
<th>Risk No.</th>
<th>Risk description</th>
<th>WP No.</th>
<th>Proposed mitigation measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Delay in ESR recruitment</td>
<td>1-4</td>
<td>A suitable amount of time for recruitment is envisioned in the network’s schedule, tailored to the different academic calendars and bureaucratic constraints known in beneficiary organisations. To soften the impact of a delayed start date of employment, the schedule leaves significant time buffers before critical milestones and events.</td>
</tr>
<tr>
<td>R2</td>
<td>ESR early dropout</td>
<td>1-4</td>
<td>The commitment of the candidates to the prospects offered by the project will be strictly assessed at the selection stage of the recruitment to minimise this risk, but the possibility of such an event due to unforeseen reasons cannot be eliminated. If it happens during the first half of the project, the ESR will be replaced following the original recruitment procedure and the new ESR will continue the project from the interruption point with complementary support from the hiring institution to complete the doctoral degree program. Otherwise, the research team of the beneficiary node will continue the project without a new hiring.</td>
</tr>
<tr>
<td>R3</td>
<td>ESR underperformanc e or long-term illness</td>
<td>1-4</td>
<td>The ESRs are the main producers of the deliverables foreseen by the program, but they are helped by a staff of participating researchers in all nodes. The latter will increase their workload to deliver the products in case the former are unable to reach the foreseen goals. The indicated delivery dates for deliverables and milestones include a contingency margin.</td>
</tr>
<tr>
<td>R4</td>
<td>Employment change or accidental unavailability of key persons</td>
<td>All</td>
<td>To account for unexpected unavailability of the network’s PIs to perform their duties, deputies will be designated in each of the participating organisations. The participation of several members per node who will cooperate in the supervision of the ESRs ensures that this risk has no bearing on that task.</td>
</tr>
<tr>
<td>R5</td>
<td>Unforeseen delay of experimental data collection</td>
<td>1 - 3</td>
<td>Although WP1, WP2, and WP3 will benefit from data that will be collected during the LHC Run 3, they are based on the large amount of data collected during Run 2 and are therefore less exposed to this risk. In the unlikely event of a very long delay in the collection of new data from the medical imaging experiments involved in WP4, the analysis techniques developed in this WP will be applied to existing data, to allow the ESRs to still complete their PhDs and produce public reports.</td>
</tr>
<tr>
<td>R6</td>
<td>Unexpected loss of analysis software and results</td>
<td>1 - 4</td>
<td>The network’s policy of wise use of code repositories on Web sites (git, rivet) and on public servers and redundant storage of all results from intermediate analysis steps will be enforced and continuously monitored throughout the research programme to ensure no loss of the scientific products. The same policy of redundant storage will also be enforced for the medical imaging data used in WP4. The LHC data that will be used in WP1, WP2, and WP3 are secured by the data preservation programs of the respective experiments.</td>
</tr>
<tr>
<td>R7</td>
<td>Unexpected cancelation of a network event</td>
<td>5 - 6</td>
<td>Although the program is designed with minimal reliance on external organisations, we believe that there is added value in the cooperation with the institutions organising schools. In case of cancelation of a school, we plan to invite the main instructors to give cycles of seminars, providing the same training to our ESRs. In case of cancelation of an outreach or dissemination workshop, we plan to reorganise it at a suitable time and member organisation.</td>
</tr>
<tr>
<td>R8</td>
<td>Conflict in the network management</td>
<td>7</td>
<td>Conflicts in the network management that could jeopardise any tasks of the project will be discussed in SB meetings and, if the need arises, resolved by majority vote. In the case of resignation of a person from a management post for any reason, the SB will select another member of the network for that post.</td>
</tr>
<tr>
<td>R9</td>
<td>Misconduct of a network member</td>
<td>All</td>
<td>We plan to impose strictly the principles and rules outlined in the European Code of Conduct for Research Integrity. All members will explicitly commit to this Code in the Consortium Agreement. In the case that a violation of these rules by a member of the network will be identified, the SB will assess the severity and intention of the misconduct and will reserve the right to expel that member from the network or, if it involves an ESR, to cancel the hiring contract and follow the steps in R2.</td>
</tr>
</tbody>
</table>

### 3.2.6 Intellectual Property Rights (IPR)

Intellectual Property Rights (IPR) will be agreed in compliance with the Commission recommendations on the management of intellectual property in knowledge transfer activities\(^\text{25}\). This will ensure that the results of the project will be available for optimal exploitation via further research activities, developing new products or creating new services. Access to the background (data, software, and computing resources) will be agreed upfront in the Consortium Agreement (CA). It will be given to the ESRs, to conduct their research activities, and to collaborating beneficiaries, to exploit their own results under fair and reasonable conditions. For the HEP-related research tasks, the data, either coming from proton-proton collisions or simulated, is property of the LHC experiments. The Open Data project guarantees access to the community at large only after a sufficiently long time. The network members of each experiment will ask the corresponding Collaboration the already in-place exceptions for external collaborators, therefore providing all network participants with the possibility to work with the most recent data. All scientific publications released by the network will have open access.

The results of the research program will be owned by the beneficiaries who create them, but access will be granted to all other beneficiaries who need them to complete their tasks and exploit their own results, under fair and reasonable conditions. Specific algorithms and methods obtained from work on commercial-facing tasks will not be shared among beneficiaries, but the general approach and technology to obtain them can be shared, so that the techniques can be applied to other fields. A joint ownership agreement for jointly generated IP will be included in the CA, dealing as well with ownership transfer details and meeting the requirements of all participants. The CA will also include terms for the beneficiaries to protect adequately the results, taking into account the interests of the other beneficiaries, if they could be reasonably exploited and the protection is reasonable. Both our non-academic beneficiaries, B12 and KROMEK, have in-house IP advisors who will participate in the CA negotiation and will contribute during the implementation of the project by actively reviewing results throughout the project, to help ensure that results having potential commercial value are protected to maximize exploitation potential. In addition, the CA will specify how the results of this project, including software products, will be used as background for future internal research conducted by the beneficiaries and collaborative research conducted by the LHC experiments.

3.2.7 Gender aspects

DESCARTES strives to achieve gender equality and to promote women in science. The gender composition of DESCARTES is similar to the world average in physics, which is unfortunately uneven\textsuperscript{26}: three out of the fourteen main supervisors are women, while in addition Barbara Alvarez (OVD) will co-supervise ESR12 (CERN). However, the women main supervisors play key roles in the network: two of them are the leaders of WP2 and WP3 and, at the same time, they will serve as SB Chair Scientific Coordinator, while the third will be the Knowledge Transfer Coordinator. We also aim to fill at least one third of the ESRs position with female students.

3.2.8 Data management plan

The Data Management Plan (DMP) of DESCARTES is the management deliverable D7.2 of WP7 (see Table 3.1b), due on month 4 of the network’s operation. It will manage the data produced by the network (including codes, publications, web content) according to a set of guiding principles outlined in Subsection 3.2.6 to make data findable, accessible, interoperable, and reusable, with emphasis to enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals\textsuperscript{27}. We remind that the experimental data collected by the LHC collaborations ATLAS, CMS, and LHCb is proprietary to CERN and will be managed according to CERN’s data preservation, reuse and open access policies.

3.3 Appropriateness of the infrastructure of the participating organisations

All the participating Institutions have established infrastructures to host researchers and provide them with appropriate means to carry out their research, as well as to help them with visa, accommodation, and other administrative issues. The ESRs will be provided with comfortable office space at their host Institutions, in close contact with their supervisors and other team members. Additional facilities, such as libraries, computing helpdesks, administrative support for handling travel for participation at workshops and conferences, or other issues will also be available.

The state-of-the-art technological infrastructure and equipment will be provided to favour training and research, at the level of particle physics facilities such as particle detectors and accelerators (a research cyclotron at UCL and, especially the LHC at CERN), laboratories for mechanical, electronic, microelectronic and optoelectronic engineering and large cryogenics installations (CERN). The HEP research work of the ESRs will happen in the context of the CERN experimental collaborations, giving them access to a wide range of expertise and all the analysis and processing infrastructure developed within the experiments. This includes, for members of the LHC Collaborations, access to resources for the production of MC simulated samples, crucial for the program of all WPs, or more specifically access to prototype FPGA boards for the upgraded CMS trigger, which are the object of part of the research program of WP4. Further examples of computing facilities, provided by all the network organizations are: the world largest MadGraph5\_aMC@NLO cluster, hosted by UCL, an infrastructure based on Cloud OpenStack middleware that connects together a pool of hybrid machines and allow GPU calculus sources (INFN and UNIPD), large centres for GRID Computing (UCL, LIP, UOXF); simulation studies based on the GEANT4/GATE environment can be performed in running GRID clusters (IASA). Additionally, two experimental, high sensitivity and resolution $\gamma$-Camera systems dedicated to SPECT imaging for medical and radio-pharmaceutical research are provided for the development of WP4(SPECT-Lab of IASA). The ESR will have access to the Hadoop and Spark services maintained by the CERN IT. These infrastructures represent an essential asset for the research and development work of WP3.

Figure 5: The HR logo.

Among the academic partners CERN, IASA, INFN, OVD, and UCL have endorsed the EURAXESS European Charter for Researchers and Code of Conduct for the Recruitment of Researchers, two key documents with which the European Union addresses the need to make research an attractive career. CERN, UOXF, UCL, and UNIPD have also been awarded the EURAXESS Human Resources Excellence in Research logo (Fig. 5), an acknowledgement of the commitment of the institutions to fair and transparent recruitment and appraisal procedures.

\textsuperscript{26} For the gender aspects in physics see the detailed survey of APS: https://www.aps.org/programs/women/resources/statistics.cfm. The situation is very similar in Europe.

\textsuperscript{27} Mark Wilkinson et al. Scientific Data 3, Article number 160018 (2016).
3.4 Competences, experience and complementarity of the participating organisations and their commitment to the programme

3.4.1 Consortium composition and exploitation of participating organisations' complementarities

The participating organisations in DESCARTES have an established record of excellence in diverse and complementary scientific domains. The network will benefit from the interdisciplinary expertise of its members to bind together different requirements coming from the tasks that the project is meant to face.

Complementarity of the participating Institutions is summarized in the following:

1. The network gathers organizations specialised in either experimental and phenomenological research in HEP (CERN, LIP, IASA, INFN, OVD, UCA, UCL, UOXF), and in inferential statistics and statistical learning methods, both at the academic level (UiO, UNIMIB, UNIPD) and at the industrial one (KROMEK, B12).
2. Organizations specialized in HEP participate to three LHC experiments - ATLAS (UOXF, UCA), CMS (IASA, LIP, UCL, OVD, CERN) and LHCb (INFN). Research experience ranges from precise particle physics measurements, and experimental test of new theoretical models, real time data selection and data analysis.
3. Organizations specialized in statistics will create a synergies by binding together advanced methods of data science (UiO, UNIPD), with the required inferential theory on which they rely (UNIMIB), and the experience of their effectiveness gained via the application to real world problems (B12).
4. All the participants have a strong interest in advanced data science tools, and have become experts in their use. Members who participated in the previous AMVA4NewPysics ITN network have gained experience with advanced ML tools (IASA, LIP, UOXF, UCL, UCA, B12) and with their application to HEP problems (UNIPD). The exploitation of this collective know-how will occur and develop thanks to frequent interactions at the network meetings.
5. The training offered to ESRs in DESCARTES will benefit from the diverse experience of the supervisors who are strongly committed to teaching and supervision.
6. The very broad offer of training detailed in Sec. 1.2 is possible thanks to the participation in the network of Institutions who bring in complementary training opportunities at the participating Institutions.
7. The industrial organizations (B12 and KROMEK) will be actively involved in research, both in data analysis and in physical applications, thus bringing to the network complementary competences to augment those in basic research at the academic Institutions.
8. Partner companies are chosen to complement the network with a wide range of external applications of ML methods, to broaden the potential of ESRs and provide them a cross training as data scientists.

3.4.2 Commitment of beneficiaries and partner organisations to the programme

The dedication and commitment of the beneficiary organisations to explore and deepen basic human knowledge is at the core of their essence and fundamental goals. Physicists at the participating Institutions devote their research time to the understanding of fundamental Science, while Statisticians to the development of the tools which make possible the advancement in any science. DESCARTES provides an excellent opportunity of further exchange of resources and ideas between researchers of different organisations, enhances interaction among the members and develops synergies among different research areas. Common deliverables will provide the opportunity to define the path of research to a clear set of goals, where the collaborative effort will play an important role. The development of ML advanced techniques is a goal common to the participants that will drive the activities to a successful project.

In particular, DESCARTES will provide an environment to promote excellence through the collaboration of researchers in different complementary projects towards the deliverables in WP1, WP2, WP3, and WP4. Researchers from the ATLAS, CMS, and LHCb experiments will have the opportunity to collaborate to the activities in which the network is involved, i.e. in new physics searches (UCA, IASA, LIP, INFN, WP1), Higgs studies (OVD, LIP, UOXF, INFN, CERN, WP2), EFT parameters (UCL, UOXF, WP3) and boost new tools relevant to the prospects of the HL-LHC (CERN, IASA, WP4). Within each WP, the development of common algorithms applied to different cases will be driven by the expertise of the organizations specialized in statistics via their involvement in each of the WPs. Consistently, the non-academic organizations designed to be beneficiary, are specialized in data science (B12) as well as physical applications (KROMEK).

Personal connection and previous collaboration are at the basis of the choice of the partners. and non-academic organisations, the latter ones committed to the main aim of enlarging the potential of ESRs and complementing their scientific training as data scientists.
### 3.5 Gantt Chart

<table>
<thead>
<tr>
<th>Months</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>A A</td>
<td>P P</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR2</td>
<td>UCA</td>
<td>P P</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
</tr>
<tr>
<td>ESR3</td>
<td>UNIPD</td>
<td>L L</td>
<td>X X</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR4</td>
<td>INFN</td>
<td>O O</td>
<td>V V</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR6</td>
<td>OVD</td>
<td>I I</td>
<td>O O</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR7</td>
<td>UiO</td>
<td>V V</td>
<td>S S</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
<td>N N</td>
</tr>
<tr>
<td>ESR8</td>
<td>UOXF</td>
<td>M M</td>
<td>C C</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR9</td>
<td>UCL</td>
<td>E E</td>
<td>M M</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR10</td>
<td>UNIMIB</td>
<td>X X</td>
<td>C C</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR11</td>
<td>B12</td>
<td>G G</td>
<td>E E</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
<td>R R</td>
</tr>
<tr>
<td>ESR12</td>
<td>CERN</td>
<td>I I</td>
<td>F F</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR13</td>
<td>IASA</td>
<td>G G</td>
<td>F F</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
<td>S S</td>
</tr>
<tr>
<td>ESR14</td>
<td>KROMEK</td>
<td>N N</td>
<td>I I</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
<td>G G</td>
</tr>
</tbody>
</table>

#### Training

- **Instit. training**
- **School ST**
- **School ML**
- **School MG**
- **Workshop MLHW**
- **Workshop SS**
- **Workshop OR**
- "Hackathon"
- **Workshop SJ**
- **Workshop KTC**
- **Dissemination**
- **Public seminars**
- **ITN blog**

#### Dissemination/public engagement

#### Management

- **Management**
- **Recruitment**
- **Events org.**
- **ITN meetings**

Meaning of letters in the chart:

- **S**=Career-development secondment. **A-X:** Research secondment (A=UCA, B=B12, C=UCL, E=ULI, F=FNAL, G=CERN, I=IASA, L=LIPI, M=UNIMIB, N=INFN, O=UiO, P=UNIPD, R=UCI, V=OVD, X=UOXF). **ST=**Statistics; **ML=**Machine learning; **SS=**Soft skills; **MG=**MadGraph; **MLhw=**ML on hardware; **SJ=**Science journalism; **OR=**Outreach.
- Lighter shading (e.g. in ESR timelines) indicates accounted contingency.

**Note:** indicated time for some secondments and events may be subject to changes.
5. Participating organisations

The tables below provide information on the beneficiary organisations.

<table>
<thead>
<tr>
<th>Beneficiary Legal Name</th>
<th>Institute of Accelerating Systems and Applications/National and Kapodistrian University of Athens (IASA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Description</strong></td>
<td>The Institute of Accelerating Systems and Applications (IASA) promotes research and postgraduate studies in the Greek university system. It is affiliated with six university departments: Medicine, Physics and Informatics of the National and Kapodistrian University of Athens (NKUA) and Electrical &amp; Computer Engineering, Chemical Engineering and General Science of the National Technical University of Athens (NTUA). IASA is an autonomous legal entity governed by a Board of Directors elected by the governing bodies of the two Universities. It is currently engaged in many cutting-edge R&amp;D projects. Several tens of graduate students (M. Sc. and Ph. D.) and postdocs have been carrying out research at IASA under the supervision of faculty from the two universities.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td>Konstantinos (Costas) Vellidis (PhD 2001) associate professor at the Physics Department of NKUA, Network Coordinator, main ESR supervisor: 60% FTE. CERN research scientist (2017-2018), CMS co-author since 2017, coordinator of the Level-1 Muon Trigger Upgrade for HL-LHC (2018-2020), participating in physics analyses involving searches for supersymmetry. Fermilab research scientist (2008-2017), co-spokesperson of the CDF Collaboration (2012-2015), convener of the CDF Top quark and BSM physics group (2011-2012), CDF Monte Carlo production coordinator (2008-2011), participated in numerous Higgs boson, top quark, and QCD physics analyses. Efstathios Stiliaris (PhD 1990) Associate Professor at the Physics Department of NKUA. His research interests are focused on the experimental Nuclear Physics and Applications, including instrumentation and development of optimization techniques for high sensitivity and resolution tomo-scintigraphic imaging devices. He holds a PhD in Experimental Nuclear Physics from the Hahn-Meitner Institute and the Freie Universitaet Berlin, Germany. He is currently acting head of the Nuclear Laboratory and the IASA SPECT-Lab at NKUA. 30% FTE.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>IASA has excellent support and stuff in all fronts: administrative, technical and computing. It provides access, via the two Universities, to essentially all scientific journals, libraries and online scientific and academic resources. IASA researchers have access, via cooperation agreements, to internationally renowned institutions. Each year IASA hosts several PhD students and Master student internships.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>IASA is housed in an independent building that belongs to NKUA. It also has extended storage space where parts of its accelerator complex are stored at the NTUA. All these premises are wholly independent from other beneficiaries and/or partner organisations in the consortium.</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes</strong></td>
<td>The Institute has had numerous successful collaborations with prominent groups in Europe and in the US. Faculty affiliated with IASA have participated in several FP5, FP6 and FP7 programmes. IASA, as a JRU member of the Hellenic National Grid Initiative (NGI-GRNET) has been involved in the EGEE-I, EGEE-II, EGEE-III and EGI-InSpire projects. It was also one of the lead partners in the FP6 GRIDCC project. In 2008, IASA developed and has since been maintaining the EGI Applications Database (<a href="https://appdb.egi.eu/">https://appdb.egi.eu/</a>). Finally, as of May 2010, the IASA GRID Operations Centre has developed and has since been operating and maintaining the backend system of the European GRID Initiative (EGI) Unified Middleware Distribution (UMD) Repository. IASA participated in the PRSATLHC RTN (ended in 2006), and was the coordinator of the FP7 project DISCOVER the COSMOS, a coordination action aiming to demonstrate innovative ways to involve teachers and students in e-Science. This action had 15 partners in the consortium.</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>IASA is currently participating in over 20 externally funded research programmes. Of them, the most relevant EU-funded projects in the thematic areas and principles of the current proposal are (1) AMVA4NewPhysics: an ETN focusing on statistical learning applications in high energy physics; and (2) GO-LAB: the project opens up online science laboratories/remote and virtual labs for large-scale use in education.</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
<tr>
<td>Beneficiary Legal Name</td>
<td>European Organization for Nuclear Research (CERN)</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td><strong>General Description</strong></td>
<td>CERN is an International European Interest Organization and is the world’s largest particle physics centre, providing technologically-advanced facilities for particle physics. CERN has 23 member states. Close to 13,000 scientists from 650 institutes worldwide are involved in the research and technology programme. CERN’s mission is focused on 4 topics: research, technology, collaboration and education, including a long and strong training tradition via the Fellows, Associates and Students programmes. It has its own Learning and Development service providing almost 14,000 person days of technical management, communication, academic, safety and language training per year.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td><strong>Gianluca Cerminara</strong>, staff physicist in the Experimental Physics (EP) Department, is PI of the CERN node and supervisor of ESR XX (15% FTE). Member of the CMS collaboration since 2002, he has been leading the CMS sub-project taking care of the offline operations of the experiment for over 6 years, serving as member of the Executive Board of the experiment. He has been driving the automation of the Data Quality Monitoring infrastructure of the experiment leading several ongoing projects for the application of deep learning techniques to the domain of anomaly detection. He is currently coordinating the group designing the algorithms for the online Level-1 trigger of the CMS experiment for the HL-LHC upgrade for what concerns the identification and selection of electromagnetic objects.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>World-class accelerator facilities: PS / SPS / LHC complexes Due to its position as a focal point for research into elementary particle physics and associated technologies, CERN is graced with state-of-the-art technological infrastructure and equipment. This spans a very large range of facilities such as accelerators and particle detectors, a forefront informatics backbone including Grid developments, state-of-the-art laboratories for mechanical, electronic, microelectronic and optoelectronic engineering and large cryogenic installations. The ESR will work in the context of the CERN group of the CMS experiment and will have access to all infrastructures provided by the laboratory and in-house engineering/technology/detector physics groups. The Knowledge Transfer group at CERN aims to engage with experts in science, technology, and industry in order to create opportunities for the transfer of CERN’s technology and know-how. The ultimate goal is to accelerate innovation and maximise the global positive impact of CERN on society. This is done by promoting and transferring the technological and human capital developed at CERN. The CERN KT group promotes CERN as a centre of technological excellence and the positive impact of fundamental research organisations on society. The Human Resources department has run many tens of EU-funded projects and training networks in the past and will take care of the administrative needs of the ESR.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>All research facilities are owned and operated by CERN. They are wholly independent from other beneficiaries and partner organisations in the consortium</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes, including ITN</strong></td>
<td>FP6: Coordinator of 7 EST, 1 RTN, 8 individual fellowships + partner in 2 RTNs FP7: Coordinator of 11 ITNs + partner in 8 others FP7: Coordinator of 4 COFUND grants FP7: Coordinator of 15 individual fellowships FP7: partner in 2 IAPPs FP7: coordinator of an Integration Grant</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>H2020: Coordinator of 5 ITNs, beneficiary in 4 others H2020: Coordinator of 2 RISE, beneficiary in 3 others H2020: Coordinator of 11 Individual Fellowships H2020: Coordinator of a COFUND grant from the 2014 Call for 60 Fellows</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
<tr>
<td>Beneficiary Legal Name</td>
<td>Università degli Studi di Padova (UNIPD)</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td><strong>General Description</strong></td>
<td>Dating back to 1222, UNIPD is one of Europe’s oldest and most prestigious seats of learning. It offers its students 32 departments, 40 doctoral degree courses activated this year, 2 international doctoral degree courses and 44 research and service centres across the spectrum of sciences, medicine, social sciences and humanities. Frontier Research@UNIPD is funded by Intramural grants (ca. 20 ME year) and European projects (ca. 15 ME year), being a first driver for technology transfer. The Department of Statistical Sciences (DIPSTAT) has a long history at UNIPD; it hosts researchers working on methodological and applied Statistics. The Department has an international research profile, as witnessed by the results obtained in the first official ranking of Italian research production (VQR 2011-2014). DIPSTAT ranked 2nd among the large departments in Area 13-Economical and Statistical Sciences. In 2019, the area “Statistics and Operational Research” of Padova has ranked among the first 100 all over the world in the QS World University Ranking. The aim of DIPSTAT is to promote knowledge and competence in Statistics at all educational levels, especially in the academic field. Ultimate excellence in education is reached within the Doctoral Programme in Statistics, offered by the Department since 1984. The Ph.D. in Statistics hosts every year about 25 Ph. D. students enrolled in the School from all over the world.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td>Giovanna Menardi, Ph.D., Associate Professor at DIPSTAT, node coordinator, and supervisor of ESRX. Her research activity focuses on classification and clustering methods in data mining and has been published in international journals of Computational Statistics, Data Mining, and Finance. She is author of software packages in R. Former PI and ESR supervisor in a former MSCA-ITN network. FTE: 40%. Davide Risso, Ph.D., Assistant Professor at DIPSTAT; co-supervisor of ESRX; Adjunct Asst. Professor, Div. Biostatistics and Epidemiology, Weill Cornell Medicine, New York. His research focuses on the statistical modeling of high-dimensional data. Author of articles in international journals and of several software packages in R and C++. FTE: 20%. Livio Finos, Ph.D., Associate Professor in Psychometry. Research participant. FTE: 10%. Federico Ferraccioli, postdoctoral fellow in statistical sciences at DIPSTAT. Research participant. FTE: 10%.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>UNIPD offers 1,761 laboratory rooms spread over all its premises, corresponding to 65,426 m². The University Library System, divided into 10 Poles, provides services through 51 libraries and widely meets the specific study and research needs of the various scientific areas. More than two million documents are available to users. Electronic resources consists of 565 databases and over 10,500 scientific journals. The University Library System belongs to local, national and international library networks. The University Language Center organizes Italian Language courses for foreign students enrolled at the University of Padova and for students and researchers coming at the University within a mobility programme. Moreover, fellows will have access to all the facilities and tools provided by the multi-media library hosted at the Language Center. Computers, software programs, and server services for computations, as well as office spaces are made available by the Department of Statistical Science.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>UNIPD owns independent research premises</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes, including ITN</strong></td>
<td>Within the FP7 framework UNIPD had 196 funded projects among which 33 were FP7-PEOPLE programme: 17 ITN, 8 Marie Curie IF, 4 RISE, 3 Researchers Night and 1 COFUND Programme. Furthermore, 16 projects were funded by the European Research Council within the specific programme “Ideas”. UNIPD also performed in a distinguished way in other EU and international programmes in the last years for a total contribution of about € 7.5 Million. DIPSTAT has been former beneficiary in a EU FP7 research and Cooperation Program (2013-2017). DIPSTAT has been former beneficiary in the AMVA4NewPhysics MSCA-ITN network (2015-2019) and 3 further EU programs.</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>UNIPD currently participates in 150 Horizon 2020 actions, and in 20 projects in other EU programmes. It is now coordinating 45 H2020 projects. Among the projects funded within Horizon 2020 there are 20 ITN, 29 IF, 10 RISE and 17 ERC. DIPSTAT is currently entitled as a Department of Excellence by the Italian Ministry of University and Research, and granted with access to special funding for a project on Statistical Models and Methods for Complex Data. It is involved in one EU research program and 2 further International research programs</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020- MSCA-ITN-2020 call</strong></td>
<td>None</td>
</tr>
<tr>
<td>Beneficiary Legal Name</td>
<td>University of Oslo (UiO)</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>General Description</strong></td>
<td>The University of Oslo (UiO) is Norway’s highest ranked institution for research and education, no. 59 on the Shanghai ranking 2019, and is among the top three most innovative universities in Scandinavia. UiO has 28,000 students, 3,000 PhD candidates and 7,000 staff whereof 3,800 full time academic and 1,120 technical support. UiO has implemented the European Charter &amp; Code for Researchers and have had the distinct HR Excellence in Research since 2009. Department of Mathematics is part of the Faculty of Mathematics and Natural Sciences and it is engaged in research covering a wide spectrum of subjects within mathematics, mechanics and statistics. The Section for Statistics and Data Science, in particular, includes 9 tenured faculty members in statistics and several postdocs and PhD students, making up a group of about 40. Statistics at UiO is internationally recognized, with interests spanning a broad range of areas (including space-time data, model selection, mathematical statistics, time-to-event models, data integration, copula and dependence models, statistical genomics, Bayesian inference, stochastic models, high dimensional data and models) and numerous collaborations with leading research groups internationally. The research group is central in the Center for Research-based Innovation 'BigInsight' which is a research partnership with the Norwegian Computing Center, other statisticians at the University of Oslo and important industrial and public partners.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td>Riccardo De Bin, Ph.D., Associate Professor at the Department of Mathematics of the University of Oslo. He is working at the interface between Statistics and Machine Learning, in the broad field of Data Science. He has strong expertise in statistical boosting, a powerful machine learning tool re-interpreted from a statistical point of view. He is also an expert in the integration of low and high-dimensional sources of information for constructing prediction models and on the use of integrated likelihoods in the frequentist framework.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure, and Equipment</strong></td>
<td>Computers, software programs, and server services for computations, as well as office spaces and library resources are made available by the Department of Mathematics.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>The University of Oslo has independent research premises.</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes, including ITN</strong></td>
<td>In FP7 (2013-2017), UiO participated in 84 projects; representing 24 ERC grants, 38 MSCA individual fellowships and 22 MSCA networks (3 as coordinator: SCIENTIA-FELLOWS (609020), ZEOMORPH (606965) and BACMT (247634)). The Faculty of Mathematics and Natural Sciences represented more than half (85) of the UiO’s EU portfolio; 55 collaborative projects (11 as coordinator), 13 ERC grants, 17 MSCA-IFs, 9 MSCA-ITNs (one as coordinator ZEOMORPH (606965)) and 1 MSCA-IRSES (one as coordinator: BACMT (247634)).</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>In H2020 (2014-2020), as of October 2019, UiO has been granted 162 projects, whereof 31 ERC grants, 64 MSCA and 67 projects within “Thematic actions”. The Faculty of Mathematics and Natural Sciences participates in 74 H2020 projects: 12 ERC grants (5 StG, 4 CoG, 1 AdG, and 2 SyG), 30 MSCA projects (13 IF, 15 ITN and 1 RISE). Our Faculty was granted two ITNs as coordinator: NanoHeal (642976) and MARmaED (675997).</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
<tr>
<td>Benefitary Legal Name</td>
<td>Université Catholique de Louvain (UCL)</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td><strong>General Description</strong></td>
<td>The Centre for Cosmology, Particle Physics and Phenomenology (CP3) at UCL consists of 12 staff members (six experimentalists and seven theorists), currently more than 25 post-docs, 25 Ph. D. students and a variety of master and bachelor students. All scientific activities are run in conjunction between the experimental and theory group. The Centre invests and fosters collaborations abroad as well as a diverse and international composition: the fraction of non-belgian researchers is always well above 50% at all levels (faculty, post-doc, Ph. D. students). The CP3 Centre at UCL has also a distinguished record in training early researchers. Only in the past five years 28 Ph. D. students have obtained their Ph. D. title with theses in high-energy physics, the majority of which have then found postdoctoral positions in the same period of time and the others have immediately been hired in the industry. More than 56 postdoctoral scientists have been educated at UCL in the last 10 years. At least eight of these have since obtained permanent or semi-permanent academic positions in other institutions.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td>Christophe Delaere (10%), node coordinator and co-supervisor of ESR9, is an experimentalist leading the effort towards the development and implementation of a new generation of Matrix Element Methods in CMS. He is also promoter of Delphes and developer of NN applications for more than 15 years. Andrea Giammanco (10%), co-supervisor of ESR 11, is a member of CMS where he coordinated the Simulations and the Top Quark groups. He is also active in muon radiography and is one of the authors of the Delphes package. Member of the Science PhD board at UCL, and Chair of the CMS Thesis Award Committee. Fabio Maltoni (10%), is Professor at UCL since 2005 and holds a joint position with Università di Bologna (IT) since Sept 2018. He is a well-known phenomenologist in the field of collider physics and main author of MadGraph5_aMC@NLO, a world-wide employed MC for the accurate and precise simulation of events at high energy. He will be the reference person for simulation tools and the related training activities.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>CP3 hosts experimental labs, clean rooms and a Tier-2 Centre for CMS, NA62 and the world largest MadGraph5_aMC@NLO cluster (&gt;500 nodes) and online servers.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>UCL owns the premises and the infrastructure where CP3 operates.</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes</strong></td>
<td>CD, AG and FM were part of the “AMVA4NewPhysics” ITN. In that context, CD was supervisor of one ESR and AG was the Deputy Coordinator and the Training &amp; Events Officer. FM has taught in all the main summer schools for Ph. D. (European (Romania, Italia), Asian Pacific (Beijing), Latin-American (Peru) CERN schools), TASI (US), Cargese (F), TAE School (Spain), INFN schools (Italy), BND schools, YETI and BUSSTEP schools (Durham, Oxford)) and he teaches regularly to the CERN summer student and academic training program. He is main organiser of the international FeynRules/MadGraph Collider Phenomenology Schools (Brazil, China, Taiwan, Japan, Korea).</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes</strong></td>
<td>AG is the coordinator of WP “Muography” of the H2020-MSCA-RISE network “INTENSE”. CD, AG and FM are involved in the “be.h” EOS program in Belgium (<a href="https://be-h.be">https://be-h.be</a>), including for all of them the supervision of PhD students and postdoctoral researchers.</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
<tr>
<td>Beneficiary Legal Name</td>
<td>B12 Consulting (B12)</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------</td>
</tr>
</tbody>
</table>

### General Description
B12 Consulting is an IT consulting firm based in Louvain-la-neuve, Belgium and adjacent to the CP3 institute. Founded in 2105, it currently consists of 18 consultants, 17 of which hold a masters’ degree in technical sciences or higher. B12 provides services in software development, advanced data analytics and strategic IT consulting to clients across a wide spectrum of business and industry.

### Role and Commitment of key persons (including supervisors)
**Michel Herquet (10%)** is a managing partner at B12 Consulting. He holds a PhD in theoretical particle physics from CP3 institute (UC Louvain). He completed one postdoc at Nikef. Michel’s scientific career was focused on phenomenological studies of new physics at the LHC as well as development of Monte Carlo tools for HEP simulations. He is one of the co-authors of MadGraph 5. As a managing partner, Michel handles day to day operations of B12 as well as business development and technical aspects of commercial projects.

**Mihailo Backovic (10%)** is a senior data scientist and a project manager at B12 Consulting. Mihailo holds a PhD in theoretical particle physics from the University of Kansas and has completed postdoctoral appointments at the Weizmann Institute and CP3 institute (UC Louvain). As a physicist, he specialised in collider and dark matter particle phenomenology, while also being involved in the development of computational tools (MadDM, TemplateTagger) and applications of deep learning in boosted jet tagging. As a data scientist, he manages all aspects of commercial analytics projects at B12.

### Key Research Facilities, Infrastructure and Equipment
B12 is currently in the process of constructing a headquarters building, in proximity of the CP3 institute which will be able to host up to 30 people, to be completed by June of 2020. B12 also hosts a dedicated server with 2 GPU units, for the purpose of deep learning projects, with a plan to introduce an additional machine of equal specifications in the near future.

### Status of Research Premises
B12 owns the building which is currently near completion.

### Previous Involvement in Research and Training Programmes
- MCNET - partner organisation (ESR secondments).
- AMVA4NewPhysics - partner organisation (ESR secondments).

### Current Involvement in Research and Training Programmes
- ITOP - Top Quark physics (beneficiary, 1 ESR, secondments)
- HIMALAYA - Explainable AI and Quantum Machine Learning in HEP (partner org., secondments)
- Best4Hep - Big Data & advanced computing methods in HEP (partner org., secondments)

### Submission of similar proposals under the same H2020-MSCA-ITN-2020 call
None.

### Relevant Publications and/or Research / Innovation Product
**Beneficiary Legal Name** | Université Clermont Auvergne (UCA)  
--- | ---
**General Description** | Université Clermont Auvergne (UCA) is a multidisciplinary higher education institution composed of 38 research laboratories working in the fields of fundamental and engineering sciences, life and health sciences as well as law, human and social sciences. UCA was created on January, 1st 2017 from the merger of Université d’Auvergne – Clermont-Ferrand I and Université Blaise Pascal – Clermont-Ferrand II. UCA will be participating in this project on behalf of its Joint Research Unit Laboratoire de Physique de Clermont (LPC) which is associated to Université Clermont Auvergne and Centre National de la Recherche Scientifique. The LPC counts about 100 staff members and 20 post-docs or Ph. D.’s. The research activities focuses on particle physics, cosmology, nuclear physics, and their applications, combining experimental results with instrumental developments and theoretical approaches. The LPC has a major involvement in the international LHC project with the ALICE, LHCb and ATLAS experiments, has a program on observational cosmology with LSST, and is involved in physics applications for health, and the environment.

**Role and Commitment of key persons (including supervisors)** | Julien Donini is professor at the UCA, he will contribute to the project at 30% FTE. Donini performs his research activities at the LPC. He has been working on experimental particle physics at the LHC in the ATLAS experiment since 2008. Before this he has worked on CMS, D0 and CDF experiments, being involved in detector R&D, software development and data analysis. His main expertise and field of interest is currently on top-quark measurements and beyond the SM searches. He was convener of the single-top group in ATLAS in 2010-2011, leading ATLAS to the first single top production measurements. In 2014 he joined the AMVA4NewPhysics ITN network as PI on the search for new physics. In parallel Donini has been leading in the past years searches for new heavy vector boson resonances (such as $W'$ boson) decaying into a top quark and a bottom quark, which resulted in several publications in peer-reviewed journals.

Vincent Barra (FTE 10%) is a full professor in Computer Sciences at UCA. He headed the engineering school in Computer Sciences, with research interests including image processing, statistical and machine learning and 3D shape analysis. He has published more than 40 papers in these fields. analysis. In 2018 he co-authored a book on machine learning entitled “Apprentissage artificiel. Deep Learning, concepts et algorithmes”. He has coordinated, and participated in, a number of national and European projects to this end.

Manon Michel (FTE 10%) is a CNRS permanent researcher since 2018 at Laboratoire de mathématiques Blaise Pascal, Université Clermont-Auvergne. She develops a line of research based on cross-fertilizing exchanges between statistical physics and computational statistics. She pioneered the development of non-reversible Monte Carlo methods based on piecewise deterministic Markov processes (PDMP) for physical systems and recent works include developments of new PDMP-MCMC solutions for Bayesian inference and computational-complexity reduction schemes.

**Key Research Facilities, Infrastructure and Equipment** | The LPC benefits from the joint facilities of both the University and the CNRS in terms of personnel, administrative and technical staff, and computing support. The LPC hosts a large Tier2 centre for GRID Computing on the French cloud. Each year the laboratory hosts several Ph. D. students and Master student internships. The laboratory will provide the ESRs all necessary infrastructures, facilities and support to ensure them a high-quality and comfortable working environment.

**Status of Research Premises** | Both the University and the CNRS own the premises and infrastructure of the LPC.

**Previous Involvement in Research and Training Programmes, including ITN** | Université Clermont Auvergne has been participating in more than 40 European projects since the beginning of the 2014-2020 European programming period. Its experience is particularly strong with Marie Sklodowska-Curie grants. LPC has been involved in 6 European projects since 2014 including 3 Marie Sklodowska-Curie Actions: AMVA4NewPhysics (ITN 2015), ELUSIVES (ITN 2015) and InvisiblePlus (RISE 2015).

**Current Involvement in Research and Training Programmes, including ITN** | UCA is currently participating in 26 European projects. 16 of them are funding under the Horizon 2020 programme. UCA is involved in 4 ongoing Innovative Training Networks: FunHoMic (ITN 2018), OrganoVIR (ITN 2018), ACHIEVE (ITN 2017) and EGRET-Plus (ITN 2015).

**Submission of similar proposals under the same H2020-MSCA-ITN-2020 call** | None.

<table>
<thead>
<tr>
<th>Beneficiary Legal Name</th>
<th>Laboratory of Instrumentation and Particles, Lisbon (LIP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Description</strong></td>
<td>LIP (<a href="http://www.lip.pt">www.lip.pt</a>) was created in 1986 as the Portuguese Laboratory for collaboration with the European Organization for Particle Physics (CERN). LIP scientific activity includes Experimental Particle Physics, Development of new Instruments and Methods, and Advanced Computing. LIP develops its R&amp;D activities mainly in the framework of experiments at CERN but also with ESA. LIP also uses other international scientific facilities such as GSI, SNOLAB, or the Pierre Auger Observatory. LIP activities also include advanced education and training, outreach, and support for education in science and technology. LIP coordinates an International Doctorate Network (IDPASC), and shares responsibilities with Portuguese Universities in joint doctoral programs. LIP has close to 200 members, including over 80 Ph.D researchers and 70 graduate students.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td>Michele Gallinaro, Researcher at LIP and professor at the University of Lisbon. He has long experience in Hadron Collider physics including the Tevatron and the LHC. At the Tevatron he contributed to the discovery of the top quark with the CDF experiment in 1995. At CMS, he coordinated the group LHC-Run1 physics analysis efforts, leading to 12 scientific publications. MG supervised several students and postdoctoral fellows. In the network he will act as supervisor investing 25% of his FTE. Joao Varela, Professor at the University of Lisbon and researcher at LIP and at CERN. Expert on detector instrumentation, SM and BSM physics. Member of the Board of the HEP Division of the EPS (European Physical Society) and editor of the Journal of Instrumentation (JINST). He participated in the creation of LIP where he served as co-director until mid-nineties. He is presently the leader of the LIP-CMS research group. In the network he will be research participants and co-supervisor with 25% FTE.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>LIP operates the LHC Tier2 GRID computing centre. It also offers to its researchers the computing resources of a Tier3 GRID centre for LHC data analysis. As part of Lisbon University, LIP gives its researchers access to the full range of University infrastructure.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>LIP has independent research premises.</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes, including ITN</strong></td>
<td>The group has 5 researchers and 5 research students and has been involved in the CMS experiment since its inception, with major responsibilities in detector design, development, commissioning, and data analysis. The LIP-CMS group had the entire responsibility of the Data Acquisition System of the Electromagnetic Calorimeter, made important contributions to the CMS Trigger System, and had major responsibilities in the Precision Proton Spectrometer (PPS) project since its inception (2014). Some of the major contributions to the CMS physics program are: (1) The discovery of a Higgs boson in the two-photons decay channel; two members of the LIP/CMS group had a major role in this analysis. (2) The study of Higgs properties: a member of the group was convener of the CMS analysis group “Higgs properties”. (3) The study of top quark production: group members performed the first measurement of the top quark mass at the LHC, and a member of the group convened the CMS “Top quark” analysis group”. (4) The search for a light charged Higgs in decays of the top quark, providing the most competitive limits. (5) Search for double Higgs production using advanced analysis techniques.</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>The group is involved in the CMS experiment and is active data analysis and detector innovation. Main areas of research are Standard Model (SM) and Beyond-SM searches (Top quark, Higgs, Dark Matter, and BSM Higgs searches), and in the detector LHC upgrade projects (PPS and Timing Detector), and medical applications (PET). Members participate in “AMVA4NP” H2020-MSCA-ITN-2015, and PICOSEC Marie Curie grants. Research members are professors at the University of Lisbon and participate in the IDPASC international doctorate program. Other FP7 projects where LIP participated are “EGI-Inspire” EINFRA-1-2014289355.; “EGI-Engage”, Call: H2020- EINFRA-2014-2 Topic: “INDIGO-DataCloud”, Call: H2020-EINFRA-2014-2, Topic: EINFRA-1-2014; FP7 Project.</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
<tr>
<td><strong>Beneficiary Legal Name</strong></td>
<td>University of Oviedo (OVD)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>General Description</strong></td>
<td>The University of Oviedo (UNIOVI) is a public institution of higher education and research in the Principality of Asturias. With over 400 years of history, it offers a full range of undergraduate degrees adapted to the European Higher Education Area (EHEA) in all branches of knowledge and postgraduate degree programmes in collaboration with national and international universities and more than 250 companies. The Department of Physics is very active in all the research fields and collaborates with major international laboratories in the world, CERN, ESRF, and others. It has around 70 staff members and participates in several academic programs that allow students to have a degree in Physics, Maths and Engineering. The department also coordinates a Master program in Advanced Physics (with two research lines: material science and particle physics and astronomy). The department also participates in several of the 24 PhD programs at the University, where students are trained to become highly qualified researchers both for an academic carrier and the private industry.</td>
</tr>
<tr>
<td><strong>Role and Commitment of key persons (including supervisors)</strong></td>
<td><strong>Javier Cuevas</strong> Full Professor at the Department of Physics. ESR supervisor: 50% FTE. Member of the DELPHI experiment at LEP, CDF at Tevatron and CMS at LHC. Deputy manager of the Spanish HEP program of the spanish agency for research (<a href="http://www.ciencia.gob.es/portal/site/MICINN/aei">http://www.ciencia.gob.es/portal/site/MICINN/aei</a>) in the periods 2014-2016 and from June 2018 to date. <strong>Santiago Folgueras</strong> (PhD 2015), Assistant Professor at the Department of Physics. ESR supervisor: 50% FTE. His research interests are focused on Higgs physics analyses as a portal for new physics, he is also involved in the Muon Trigger Upgrade for the HL-LHC. Convenor of the CMS Muon physics group (2017-2019). CMS co-author since 2009, where he has participated in numerous Higgs Boson, top quark and BSM physics analyses.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>The Department of Physics is hosted in the Faculty of Sciences at UNIOVI, and has several offices and labs in other buildings spread across the University Campus. The High Energy physics group currently hosts a computing center that provides support and computing resources for the needs of the group and their research within the CMS experiment at LHC.</td>
</tr>
<tr>
<td><strong>Status of Research Premises</strong></td>
<td>University of Oviedo’s Physics department has independent research premises dedicated mainly to HEP, but also to Solid State Physics, Astrophysics and Applied Physics.</td>
</tr>
<tr>
<td><strong>Previous Involvement in Research and Training Programmes, including ITN</strong></td>
<td>UNIOVI has previously involved in the following EU-funded training programs: FP7-PEOPLE-2013-CIG, H2020-MSCA-ITN-2015-ETN, H2020-MSCA-ITN-2014-EID. UNIOVI High Energy Physics group has had numerous successful collaborations with prominent groups in Europe and in the major international laboratories, mainly around accelerators at the energy frontier, like LEP and LHC at CERN and Tevatron at Fermilab.</td>
</tr>
<tr>
<td><strong>Current Involvement in Research and Training Programmes, including ITN</strong></td>
<td>UNIOVI currently manages about 90 publicly funded projects, 10% of which are EU-funded and over 460 research contracts with companies. Staff of the Physics department has been awarded with one ERC Starting grant. UNIOVI is currently involved in: H2020-MSCA-IF-EF-ST-2016, H2020-MSCA-IF-2018, H2020-MSCA-ITN-2018, MCSA-H2020-ITN-2019. The High Energy physics group is currently participating in several externally funded research programmes either regional, national or european. Highlighting the EU-funded projects: the group is currently involved in a COST action (VBSCan) and was involved in AMVA4NewPhysics: an ETN focusing on statistical learning applications in high energy physics.</td>
</tr>
<tr>
<td><strong>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</strong></td>
<td>None.</td>
</tr>
</tbody>
</table>
### Beneficiary

<table>
<thead>
<tr>
<th>Legal Name</th>
<th>Università degli studi di Milano-Bicocca (UNIMIB)</th>
</tr>
</thead>
</table>

### General Description

The University of Milano-Bicocca (UNIMIB), founded in 1998, is one of the most dynamic, research- and innovation-oriented Italian universities, inserted in relevant networks with top universities, research centres and corporations. It is a well-organised complex, having invested in teaching support services, state-of-the-art laboratories and modern infrastructures. UNIMIB includes 14 Departments with 33,000 enrolled students, out of which 2400 international ones. UNIMIB has 117 patents and 13 spin-offs, exploiting opportunities from a broad range of technological commercial sectors. According to the National Agency for the Rating of University Quality and Research (ANVUR), UNIMIB is the 2nd highest ranked Italian university among those comparable in size. UNIMIB has extensive experience in the managing EU grants, being currently involved in 84 EU-funded projects, out of which 66 are under H2020 (for a total of 24M€). In several projects UNIMIB acts as coordinating institute.

The Department of Economics, Management and Statistics (DEMS) was founded in 2012 and counts about 73 permanent staff, several postdocs, and PhDs. Thanks to the outstanding results obtained by the department in the quality of research assessment (VQR), in January 2018 the DEMS has been awarded with the Department of Excellence 2018-2022 grant of the Ministry of Education, University and Research (MIUR). DEMS hosts the new PhD in Economics and Statistics (ECOSTAT), which is particularly active in collaborating with prestigious foreign universities, in terms of both students and faculty members exchange programs and joint degrees.

Nicola Lunardon, Ph.D., Associate Professor at DEMS, node coordinator, and supervisor of ESRX. His research activity focuses on statistical theory for pseudo-likelihood functions, resampling algorithms, and has been published in international journals of Theoretical and Computational Statistics, and Medicine. He is author and coauthor of software packages in R. FTE: 40%. Aldo Solari, Ph.D., Associate Professor at DEMS; co-supervisor of ESRX. His research focuses on high dimensional data, model selection, multiple testing, and selective inference and has been published in international journals of Theoretical and Computational Statistics, and neuroscience. Since 2014 he is a member of NeuroMi (Milan Center for Neuroscience). FTE: 30%.

### Key Research Facilities, Infrastructure and Equipment

UNIMIB possesses a well equipped environment in terms of logistic support, including office space, rooms, videoconference systems, student housing support offices, scientific library and University Network with access to all relevant publications. UNIMIB is active in a wide array of research topics in the fields of Medicine, Economics, Education, Law, Mathematics, Physics, Psychology and Sociology, thanks to its 70 Research and Excellence Centres connected with the CNR (www.CNR.it). These structures cooperate together, with different consortia, and with public and private third parties from outside the University. Centres of Excellence, Research centres and Consortia have been created to manage large multidisciplinary research activities and are managed according to guidelines that guarantee high scientific standards. UNIMIB has invested heavily in technologies to enhance and strengthen learning effectiveness. For this reason, a Multimedia Production Centre (http://www.cpm.unimib.it) has been constituted with the aim of developing a wide range of e-learning tools.

### Status of Research Premises

All research premises are owned by UNIMIB and wholly independent from other beneficiaries and partners.

### Previous Involvement in Research and Training Programmes, including ITN

Previous MSCA-ITN grants within H2020: NABBA, GA#642028 (coordinator); TOLLerant GA#642157 (coordinator); EpiPredict GA#642691 (beneficiary), ALKATRAS GA#675712 (beneficiary).

### Current Involvement in Research and Training Programmes, including ITN

Current MSCA grants within H2020: UNIMIB hosts 3 MSCA-IF (HTEPV #745304, FoodLoss #797802, READIT #792849); is coordinator of 4 MSCA-ITN (4PHOTON #721394, TRuStEE #721995, ARCH #813091, RENOIR #860034) and beneficiary in 5 more ITN actions (YEASTDOC #764927, MOTION #765298, MultiMind #765556, BactiVax #860325, DoSSIER #860721); and is coordinator of two MSCA-RISE (IPaDEGAN #778010, PANGAIA #872539).

### Submission of similar proposals under the same H2020-MSCA-ITN-2020 call

None.

### Relevant Publications and/or Research/Innovation Product

<table>
<thead>
<tr>
<th>Beneficiary Legal Name</th>
<th>INFN - Sezione di Padova (INFN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Description</td>
<td>INFN – Padova is one of the divisions of the Italian National Institute for Nuclear Physics research (<a href="http://www.infn.it">www.infn.it</a>). The INFN is the Italian research agency dedicated to the study of the fundamental constituents of matter and the laws that govern them. It operates under the supervision of the Italian Ministry of Education, Universities and Research (MIUR). It is constituted by 20 divisions and four national laboratories with the headquarters located in Rome. Today INFN employs about 2,000 scientists and supports the research activities of about 3,000 associates whose work is recognised internationally. Theoretical and experimental researches are conducted in the fields of sub-nuclear, nuclear, and astroparticle physics. INFN research areas requires the use of cutting-edge technologies and instrumentation, that are developed in the Institute laboratories and in collaboration with the world of industry. Physics and technology researches are conducted in close collaboration with the academic institutions in Italy and in the World. The Padova division includes several groups of researchers participating in over a dozen international collaborations in nuclear and particle physics; among them the LHCb-Padova group, the computing and the recent born quantum computing group to which the network participants belong.</td>
</tr>
<tr>
<td>Role and Commitment of key persons (including supervisors)</td>
<td>Prof. Donatella Lucchesi is the node PI and supervisor of ESR6. She is full professor at the department of Physics and Astronomy of Padova and she has a research appointment at INFN. Donatella is a member of the CDF collaboration where coordinated the software and computing. She was part of the group that designed the first b-jet trigger at hadron collider. Donatella is a member of the LHCb-Padova group and participates to the b- and c-jets identification using advanced software tools. Dr. Lorenzo Sestini is a researcher at INFN - Padova division. He is working on jet identification and Higgs searches with the LHCb experiment. He currently coordinates the jet reconstruction sub-group in LHCb and he is involved on the searches for H→bb/ccc with advanced statistical techniques. Dr. Alessio Gianelle is an INFN computing expert. He has a laurea degree in Mathematics, since several years works on developing software previously for the GRID and recently he contributes to the LHCb code for jets identification by using advanced methods, including quantum algorithms. Dr. Davide Zuliani is a PhD student at the Department of Physics and Astronomy of the Padova University and he is associated researcher at INFN. He is collaborating with the LHCb experiment on jet identification by using advanced algorithms including quantum information minimization algorithms.</td>
</tr>
<tr>
<td>Key Research Facilities, Infrastructure and Equipment</td>
<td>The INFN Padova division is hosted in the building of the Physics and Astronomy department “Galileo Galilei” of the University of Padova. INFN-Padova and Physics and Astronomy department share seminar and conference rooms, electronic and mechanical shops and any kind of laboratories. INFN-Padova, the Physics and Astronomy department together with one of INFN laboratories in Legnaro, located less than 10km southeast, host a computing centre which is a WLCG Tier-2. INFN-Padova was one of the proponents of the CloudVeneto, a center of competence and of excellence which serves several disciplines.</td>
</tr>
<tr>
<td>Status of Research Premises</td>
<td>The terms between INFN and the Department of Physics and Astronomy are ruled by an Agreement which easily allow, for example, the enrollment of students in the department PhD school.</td>
</tr>
<tr>
<td>Previous Involvement in Research and Training Programmes, including ITN</td>
<td>Donatella Lucchesi has participated to FP7-PEOPLE-IOF-2008 with ITES project and to H2020 with EOSCpilot for H2020-INFRADEV-2016-2 and HNsciCloud for H2020-ICT-2015 projects. INFN-Padova has lead AMVA4NewPhysics for H2020-MSCA-ITN. In addition INFN has participated to several projects in formation and research in the past.</td>
</tr>
<tr>
<td>Current Involvement in Research and Training Programmes, including ITN</td>
<td>INFN is currently involved in a number of projects related to research and training. Among the most relevant for this proposal it is worth to mention the Fellini project. Donatella Lucchesi is participating to EOSC-Pillar in response to H2020-INFRAEOSC-2018-3.</td>
</tr>
<tr>
<td>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</td>
<td>None.</td>
</tr>
<tr>
<td>Role and Commitment of key persons (including supervisors)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Prof. Daniela Bortoletto is the node PI and supervisor of ESR 8. The focus of her research is the study of the Higgs boson. She participated in the hunt for Higgs boson at the Tevatron and in its discovery at the LHC by studying the $H\rightarrow ZZ$ and $H\rightarrow bb$ decay modes. She has received numerous awards (including a U.S. NSF Early Career Award and an A.P. Sloan Fellowship) and is a Fellow of the APS, AAAS and the IOP. She has authored over 1,800 papers. She is a sought after lecturer in many programs including the CERN summer School and CERN academic training. She has been a member of the U.S. Particle Physics Projects Prioritization Panel (P5), the HEP Advisory Panel to the U.S. DOE and NSF, and the chair of the Fermilab Program Advisory Committee. She is currently a member of the KM3NET STAC and the ICFA Advisory Committee. She leads with Hayes and Shipsey the UOXF ATLAS Higgs Physics group. Bortoletto will dedicate 10% of her time to supervising ESR 8 and ITN administration, and in addition she will dedicate significant research time to WP3.</td>
<td></td>
</tr>
<tr>
<td>Prof. Chris Hays is a well-known expert in Electro-Wick, Higgs physics, and EFT methods. He is a member of the ATLAS collaboration and contributed to the 2012 discovery of the Higgs boson using its decay to a pair of W bosons. He will contribute both to the ESR training and research.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key Research Facilities, Infrastructure and Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The six-story Denys Wilkinson Lab hosts offices, a common room, seminar rooms, experimental labs, and computing facilities including two computational clusters. The primary cluster, used by the local physicists, has a size and capacity driven by activities in the two major LHC experiments, ATLAS and LHCb. Additional capacity covers the neutrino and smaller experiments and the John Adams Accelerator Institute needs. This cluster currently has a total of 700 logical CPU cores and 1300 TB of disk and provides a well-maintained environment for interactive analysis and development. UOXF also runs a Tier-2 as a part of the UK GridPP project that provides the UK contribution to the WLCG (World-wide LHC Computing GRID). The cluster is part of the distributed SouthGRID project and currently has a capacity of 3000 logical CPU cores with 950 TB of storage. 46% of the CPU in 2018 was used by ATLAS and ~90% of the storage is for their use. Other LHC experiments used a further 37% of the CPU. We have a dedicated 10Gbit/s link out to JANET specifically for the GridPP cluster’s use.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Status of Research Premises</th>
</tr>
</thead>
<tbody>
<tr>
<td>All research facilities are owned by UOXF and wholly independent.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Previous Involvement in Research and Training Programmes, including ITN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford was one of the largest recipients of FP7 funding across the EU and in FP7 was awarded &gt;200 individual Marie Curie Fellowships and 53 ITN projects (including 8 in the Physics Dept), and 39 ITNs under H2020. Oxford has 91 H2020 ERC awards to date and coordinates 10 collaborative Research and Innovation projects. UOXF also has 5 MSCA RISE projects and held a large MSCA Researcher’s Night event in September 2017.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current Involvement in Research and Training Programmes, including ITN</th>
</tr>
</thead>
<tbody>
<tr>
<td>At any given time UOXF is supporting more than 70 Marie Curie research fellows, and currently has 26 MSCA ITNs on-going. Bortoletto was a member of AMVA4Newphysics and an associated member in the FP7 INFIERI ITN.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Submission of similar proposals under the same H2020-MSCA-ITN-2020 call</th>
</tr>
</thead>
<tbody>
<tr>
<td>None.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevant Publications and/or Research / Innovation Product</th>
</tr>
</thead>
</table>
### Beneficiary

<table>
<thead>
<tr>
<th>Legal Name</th>
<th>Kromek Ltd (KROMEK)</th>
</tr>
</thead>
</table>

#### General Description

Kromek Ltd is part of the Kromek Group PLC, the world’s leading supplier of cadmium zinc telluride (CZT) radiation detectors. Formed in 2003 to commercialise crystal growth technology at the University of Durham, Kromek has grown to become a leading player in end-user and original equipment manufacturer (OEM) supplies of radiation detection in the medical, nuclear and security markets. Notable products include Kromek’s Identifier, the first equipment to be certified to the highest standard to liquid explosives detection for aviation detection, the D3S combined gamma and neutron handheld detector developed for the US Government, and a whole body gamma camera for nuclear medicine.

#### Role and Commitment of Key Persons (including supervisors)

<table>
<thead>
<tr>
<th>Key Persons</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Ben Cantwell</td>
<td>Kromek’s Innovation Director and supervisor of ESR 14. His role at Kromek includes the identification of innovation inside and outside Kromek, and fostering the development of techniques. This includes building networks with academia and industry, sourcing funding for R&amp;D projects, and identifying and capturing Intellectual Property.</td>
</tr>
<tr>
<td>Mr Ian Radley</td>
<td>Kromek’s CTO, with over 35 years’ experience of technology development. He oversees all of Kromek’s technology development, and is committed to exploring avenues for developing and adopting AI into Kromek’s product portfolio.</td>
</tr>
</tbody>
</table>

#### Key Research Facilities, Infrastructure and Equipment

Kromek is a vertically integrated company, with capabilities throughout the product development and production chain, from crystal growth, detector fabrication, electronics and applications design and simulation, through to end-user product design and build. These facilities include extensive material and detector characterization techniques, alongside detector and electronics modelling, which can be used for research, development and production. Kromek’s headquarters are in Sedgefield, County Durham, which house most of the research and production facilities outlined above. The company also has a smaller research facility in Huddersfield, Yorkshire, which specialises in neutron and scintillator detector technologies. The Group also has production facilities in Pennsylvania, USA, along with an electronics design facility in California.

#### Status of Research Premises

All facilities are owned by Kromek and wholly independent.

#### Previous Involvement in Research and Training Programmes, including ITN

Kromek has received funding for research and development from a wide range of programmes, including H2020, Eurostars, InnovateUK and various US Government programmes. It has not participated in any formal Training Programmes, although it has close links with several UK universities with whom it has arranged a variety of training links, including internships and group projects.

#### Current Involvement in Research and Training Programmes, including ITN

Kromek is currently carrying out a Eurostars project for the development of detectors for breast cancer imaging with the Technical University of Denmark. It is leading a research project into event-based processing for radiation events, funded by the US Defense Threat Reduction Agency, alongside Edinburgh, Liverpool and Manchester Universities. This project includes the training of 4 Ph.D. students.

#### Submission of similar proposals under the same H2020-MSCA-ITN-2020 call

None

#### Relevant Publications and/or Research / Innovation Product

The most relevant Innovation Products can be found on the Kromek website, www.kromek.com

- **Identifier** liquid explosive detector; an X-ray based airport scanner which uses the signal from a semiconductor detector to identify explosives in a bottle.
- **D-Matrix** CZT gamma detector for SPECT imaging; this technology has been developed to allow use across nuclear imaging modalities, including breast, heart and whole-body imaging.
- **D3S** wearable personal radiation detector; developed with funding from the US Defense Advanced Research Projects Agency (DARPA), this device uses detector gamma ray and neutron spectra to identify threat isotopes across a city.
The tables below provide information on the partner organisations.

<table>
<thead>
<tr>
<th>Partner Organisation Legal Name</th>
<th>Fermi National Accelerator Laboratory (FNAL)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General description</strong></td>
<td>Fermi National Accelerator Laboratory (Fermilab) operates under the United States Department of Energy. It was founded 1967 and has since then focused entirely on particle physics. It has approximately 2000 employees, including scientists, engineers, technicians and many thousand users from around the globe. It was the host for the Tevatron, at the time the highest energy particle collider. Fermilab is currently building new generation neutrino experiments, partnering with CERN on the LHC accelerator and physics, and has a vibrant astrophysics and cosmology program.</td>
</tr>
<tr>
<td><strong>Key Persons and Expertise</strong></td>
<td>Sergio Jindariani, PhD 2007, Co-coordinator of the LHC Physics Center (LPC) at Fermilab. Extensive experience with large scale data analysis, real-time processing, advanced algorithm development and computing architectures. Many years of experience working with FPGA and ASIC devices for high speed pattern recognition. Among the original developers of his4ml – framework for deploying machine learning based algorithms on FPGA platforms. Organiser of many international events, workshops, schools etc.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>Fermilab possesses unique computing facilities, involving local computing CPU clusters and access to GPU and FPGA devices. Fermilab also possesses unique scientific and engineering expertise in high speed data processing and machine learning. Open source software and firmware tools available at Fermilab allow for short development cycles and steep learning curve for non-experts.</td>
</tr>
<tr>
<td><strong>Previous and Current Involvement in Research and Training Programmes</strong></td>
<td>Fermilab is hosting and involved in dozens of different particle physics experiments, with many doctorate students graduating every year. In particular, the LPC hosts approximately 20 graduate student on permanent basis with many more visiting every year. Annual educational Summer programs bring tens of undergraduate students to the lab for 10-week long internships. Jindariani served as an advisor for many of them. He also organized many research and education events at the lab, including Hadron Collider Physics Summer School (80 students) and Annual CMS Data Analysis School (60-70 students each year).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partner Organisation Legal Name</th>
<th>University of California, Irvine, Department of Physics and Astronomy (UCI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General description</strong></td>
<td>Faculty focused on research in physics, with a large group of particle experimental faculty as well as world-class particle theory researchers.</td>
</tr>
<tr>
<td><strong>Key Persons and Expertise</strong></td>
<td>Daniel Whiteson: Full professor, with two decades of research experience in particle physics, and a long track record of publication in machine learning and cross-field collaborations. In addition, a strong Machine Learning group in the computer science department, led by Pierre Baldi.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>Office space, powerful computing clusters, excellent weather.</td>
</tr>
<tr>
<td><strong>Previous and Current Involvement in Research and Training Programmes</strong></td>
<td>MAPS-NSF: a graduate student training program which builds connections between machine learning and the physical sciences and fosters a local community for positive intellectual interactions. AMVA4NewPhysics: UC Irvine was a site for secondments of young researchers in this network. Several students came and spent time in our group and had very positive experiences.</td>
</tr>
<tr>
<td>Partner Organisation Legal Name</td>
<td>University of Liège, Department of Electrical Engineering and Computer Science (ULiège) (known as the Montefiore Institute)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>General description</strong></td>
<td>The Montefiore Institute is part of the School of Applied Sciences, specialized in electrical engineering, electronics, systems and modeling, optimization, machine learning and artificial intelligence.</td>
</tr>
<tr>
<td><strong>Key Persons and Expertise</strong></td>
<td>Gilles Louppe: Associate Professor in artificial intelligence and deep learning. Analysis Consultant and Expert (ACE) with the ATLAS experiment at CERN. Research interests: simulator-based likelihood-free inference, deep generative models, probabilistic programming, and developments towards the automation of science.</td>
</tr>
<tr>
<td><strong>Key Research Facilities, Infrastructure and Equipment</strong></td>
<td>Access to a national-wide consortium of computing clusters, totaling more than 10000 nodes and mixing both CPUs and GPUs with large amounts of memory and disk space.</td>
</tr>
<tr>
<td><strong>Previous and Current Involvement in Research and Training Programmes</strong></td>
<td>Previous involvement: Within FP7, ULiège was the beneficiary of 32 Marie-Curie projects, in which 9 ITN, 9 individual fellowships and 1 COFUND project for Post-Doctoral training (600405-BeIPD). Current involvement: Within H2020, ULiège is the beneficiary of 15 Marie-Curie projects, in which 8 ITN, 4 individual fellowships and 1 COFUND project for Post-Doctoral training (600405-BeIPD, FP7 but still ongoing).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partner Organisation Legal Name</th>
<th>AI4B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Description</strong></td>
<td>AI4B is a consultancy specialised in providing private sector companies and non-profit organisations a business development service based on the utilisation of novel data analysis techniques. As a service provider of data science solutions it specialises in delivering beginning-to-end software implementations specific to the needs of the client. AI4B offers training courses in areas relating to data science, thereby supporting the development of in-house expertise in machine learning and data analysis.</td>
</tr>
<tr>
<td><strong>Key Persons and Expertise</strong></td>
<td>Michael Spannowsky earned his PhD in particle physics from the University of Munich in 2007 with a thesis on Higgs boson phenomenology. He has a very broad background in particle phenomenology, ranging from precision calculations, over Higgs phenomenology to the development of analysis strategies for new physics searches. He is an expert in using novel data analysis and hypothesis testing methods, e.g. the so-called matrix element method, neural networks or boosted decision trees. He is co-founder of AI4B where he and his team provide data-science solutions to private sector companies.</td>
</tr>
<tr>
<td><strong>Key research facilities, infrastructure and equipment</strong></td>
<td>AI4B is located in Durham, County Durham, UK. Due to its close proximity of the Institute for Particle Physics Phenomenology at the University of Durham and the close collaboration between AI4B and the IPPP, the ESR will automatically obtain a visitor status at the IPPP during his/her tenure at AI4B. This will allow the ESR to attend training courses, workshops and lectures offered by the IPPP. The IPPP is the largest university-based institute for particle physics phenomenology in the world and has a strong research focus on machine learning and data analysis techniques.</td>
</tr>
<tr>
<td><strong>Previous and Current Involvement in Research and Training Programmes</strong></td>
<td>Michael Spannowsky is full professor at the University of Durham and director of the Institute for Particle Physics Phenomenology (IPPP). As PI of the IPPP grant and head of the Particle Theory section at the University of Durham, he has personnel responsibility for 90 staff members. Academically, he has supervised 6 PhD students and 19 MSc students on various research topics to completion.</td>
</tr>
<tr>
<td>Partner Organisation Legal Name</td>
<td>General Description</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Achilles Information Limited (ACH)</td>
<td>Achilles provides supply chain risk management solutions for some of the largest organisations in the planet. It collects and analyse data on hundreds of thousands of companies world-wide, through companies themselves, private organisations such as credit agencies, or publicly available datasets (world wide web/governments databases). Data collected covers key aspects of organisations risks: ability to deliver, financials, health and safety, corporate social responsibility. The Achilles Analytics department develops innovative tools that allow for near-real time risk management in these key areas using the latest advancements of statistical and machine learning tools. Achilles goal is to increase legal compliance and social accountability of companies worldwide.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key Persons and Expertise</th>
<th>Key research facilities, infrastructure and equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabrizio Margaroli earned his PhD in particle physics from University of Bologna in 2007 with a thesis on using neural networks to isolate top quark production at the CDF experiment at Fermilab, and measure its properties He later worked at Purdue University/Fermilab, Sapienza University/CERN applying advanced data analysis techniques to advance knowledge of the top quark and Higgs boson, and to search for new physics. He also served leading positions at the CDF and CMS experiments, and organised several international workshops and conferences. At Achilles he is Lead Data Scientist managing a team of 6 Data Scientists and Software Developers leveraging statistics and machine learning to create innovative products.</td>
<td>The Analytics department is located at the company headquarters in Abingdon, Oxfordshire, UK. The personal computing facilities are available in the main office, while dedicated high-performance servers are located in London and accessible remotely. Achilles partners with key cloud computing providers for additional computing needs.</td>
</tr>
<tr>
<td>Simone Gelli earned his PhD in particle physics at Sapienza University in 2017, working at the CMS experiment at CERN. There he developed a broad experience in data analysis, hardware and software development. At Achilles, he leads several projects for automated analytics.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevant Publications and/or Research/Innovation Product</th>
<th>Previous and Current Involvement in Research and Training Programmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Search for the associated production of the Higgs boson with a top-quark pair”; CMS collaboration JHEP 1409 (204) 087. “Direct constraints on the top-Higgs coupling from 8TeV LHC data” S.Biswas, E.Gabrielli, F.Margaroli, B.Mele JHEP 1307 (2013) 073.</td>
<td>Fabrizio Margaroli served during his academic experience as mentor, supervisor or co-supervisor for 1 BS, 3 MS, 4 PhD students, and one post-doctoral researcher. He now leads a team of five junior and one senior data scientists.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partner Organisation Legal Name</th>
<th>General Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWave S.L. (TW)</td>
<td>TWave is a company specialized in the design and manufacture of supervising and monitoring systems for industrial machinery. We develop innovative solutions that integrate the most advanced technologies, helping our clients to protect their critical assets through online diagnosis. In this sense, our products are designed for Industry 4.0, applying the Internet of Things to the industrial sphere. Predictive maintenance by vibration analysis is the main area of application of TWave devices, whose versatility makes them compatible with other predictive techniques such as ultrasonic analysis. The TWave technical team has been developing vibration monitoring systems for ten years. Hundreds of units have been successfully installed in industrial sectors such as wind, chemical, petrochemical or solar-thermal.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key Persons and Expertise</th>
<th>Key Research Facilities, Infrastructure and Equipment</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Previous and Current Involvement in Research and Training Programmes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation in regional and national R&amp;D programs (RIS3-Asturias Enterprise Program). The promoters have previous experience of participation in the FP7 of the EU</td>
<td></td>
</tr>
</tbody>
</table>
Relevant Publications and/or Research / Innovation Product

Several collaborations between TWave promoters and the University of Oviedo for electronic control systems and electronic instrumentation seminars and magazines, both at national and international levels.

<table>
<thead>
<tr>
<th>Partner Organisation Legal Name</th>
<th>Artabro Tech S.L. (ART)</th>
</tr>
</thead>
</table>

**General description**
Artabro Tech is a data science consulting and professional services firm operating on the highest end of the technology development. It specializes in data engineering, machine learning, artificial intelligence, and big data. The company’s core business is related to Industry 4.0 and Business intelligence services.

**Key Persons and Expertise**
- **Alejandro Casteleiro** (CEO & Co-Founder). PhD in telecommunications engineering. Over 20 years’ experience in domestic and international (Google Zurich) software development projects and in the strategic management of telecommunications operators.
- **Miguel Vidal** (CTO & Co-Founder). PhD in particle physics. Over 10 years’ experience in data analysis, machine learning and group management in international Labs (CERN and Fermilab).

**Key Research Facilities, Infrastructure and Equipment**
Office space, powerful computing resources.

**Previous and Current Involvement in Research and Training Programmes**
Artabro Tech, a two-year old startup, has no previous experience in training programmes. However, one of the promoters has large experience in research programmes, including AMVA4NewPhysics a Marie Skłodowska-Curie Innovative Training Network.

**Relevant Publications and/or Research / Innovation Product**
As a startup, Artabro Tech has no published work yet.

### 6. Ethics issues

The main objective of the research planned in DESCARTES is the furthering of human knowledge in fundamental physics, which does not involve any apparent ethical issues connected to human foetuses, human beings, human tissues, personal data, animals, third countries, environment and health and safety, nor any military applications. Similarly, the methodology of the planned research work does not include in any way organic matter, living beings, sensitive data which may involve privacy issues, or other material or topics which raise ethical issues. We do not foresee any military application of any of the products we plan to develop.

We abide and respect all fundamental ethics principles reflected in the Charter of Fundamental Rights of the European Union.

We responsibly confirm that we identify no ethics issues for which a self-assessment is needed; accordingly, no issues have been checked with a “yes” in Sec. 4 of Part A of the proposal.

### 7. Letters of commitment

Letters of commitment from the academic (FNAL, UCI, ULI) and non-academic partners (AI4B, ACH, TW, ART) are attached below.
12/27/2019

To: Prof. Konstantinos Vellidis, network coordinator
Institute of Accelerating Systems and Applications
University Campus, Physics Dept., Building IX, Zografiou - Greece
P.O. Box 17214, Athens 10024, Greece

Dear Prof. Vellidis,

I am writing this letter to express my support for the DESCARTES (aDvanceEd StatistiCal Algorithms for paRTicleE phySICS) project, which aims to train early-stage researchers (ESRs) in the areas of advanced statistical learning and novel computing techniques needed for the efficient analysis of the large and complex multidimensional data samples from particle physics experiments.

In particular Machine Learning (ML) algorithms and applications have been quickly expanding in science and industry applications. ML methods have also been proven to be very powerful in Large Hadron Collider physics, and particle physics as a whole. However, exploration of the use of such techniques in low-latency, low-power FPGA (Field Programmable Gate Array) hardware has only just begun. FPGA-based trigger and data acquisition systems have extremely low, sub-microsecond latency requirements that are unique to particle physics. Fermilab is leading efforts in developing solutions for deploying ML algorithms in FPGA based hardware. We recently developed a compiler package based on High-Level Synthesis (HLS) called hls4ml to deploy machine learning models on FPGAs.

We are willing to co-supervise the ESR’s and provide necessary education and training for them to become proficient in creating advanced ML algorithms and implementing them in FPGA-based hardware. We will share with them computing and hardware resources, software tools and techniques, as well as example ML models during their visits to Fermilab. We will also be happy to assist the project in identifying potential applications for the techniques developed by the ESRs. In addition, we will plan to organize a workshop in approximately two years from now for the ESR’s to attend and learn about leading developments in the areas of ML and advanced computing.

I believe the Fermilab LHC Physics Center and the DESCARTES project have a lot of common and synergistic research ideas and I would like to see this collaboration to strengthen and grow in the years to come.

Sincerely,
Sergo R. Jindariani
LHC Physics Center Co-Coordinator
Scientist, Fermi National Accelerator Laboratory
Prof Konstantinos Vellidis, network coordinator
Institute of Accelerating Systems and Applications
University Campus, Physics Dept. Building IX
Zografou, Greece
PO Box 17214 Athens 10024, Greece
konstantinos.vellidis@cern.ch

Letter of Commitment

Dear Dr. Vellidis,

I would like to express our full support to the aDvanceEd StatistiCal Algorithms for paRTicE physIcs (DESCARTES) project, which will provide a framework for training early-stage researchers on the advanced statistical learning and novel computing techniques needed for the analysis of large and complex multidimensional data samples, such as those acquired in the European Research Organisation for particle physics (CERN) experiments.

At UC Irvine, we have a vigorous effort to apply machine learning strategies to the challenges of high energy physics and are very interested to collaborate with your network.

We can offer early-stage researchers a stimulating environment of physicists and computer scientists, as well as a larger community that bridges machine learning and physical sciences (http://maps-nsf.uci.edu). In that context, we are very happy to host trainees from your network, as we hope their diverse experience and backgrounds will contribute to the richness of our community. The evolution of the trainees’ research will be supervised by our personnel during the time they’ll be working in our office.

Best regards,

Daniel Whiteson
Professor
Physics & Astronomy
UC Irvine
daniel@uci.edu
Dec 19 2019
December 27, 2019

Prof. Gilles Louppe  
g.louppe@uliege.be  
Department of Electrical Engineering and Computer Science  
School of Engineering  
University of Liège  

Quartier Polytech 1  
10, Allée de la découverte  
4000 Liège, Belgium

Dear Prof. Konstantinos Vellidis,

I, undersigned Gilles Louppe, in my quality of Associate Professor in Computer Science of the University of Liège, would like to express my full support to the DESCARTES project proposal submitted within the call H2020-MSCA-ITN-2020.

On behalf of the University of Liège, I also confirm that we will participate and contribute to the research, innovation and training activities as planned in this project. In particular, the University of Liège will be involved in the secondment of a PhD student under the supervision of Prof. Andrea Giannino (University of Louvain-La-Neuve, geographically close to Liège). The research project will explore the application of likelihood-free inference algorithms to high-energy physics.

I hereby declare that I am entitled to commit into this process the entity I represent.

With my best regards,

Prof. Gilles Louppe
Dear Prof. Vellidis,

It is my pleasure to support the application of DESCARTES for an ETN, which aims to train early-stage researchers (ESRs) in novel statistical methods and data science techniques. AI4B considers these methods of fundamental importance for science and private sector applications.

As an IT consultancy, specialized in providing private sector companies and non-profit organisations a business development service based on novel data analysis techniques, AI4B has strong expertise in various machine learning techniques and statistical methods. We are fully committed to co-supervise an ESR and are available to host 2-3 secondments. During the secondments of the ESR to AI4B, we will provide him/her with education and training opportunities in novel data science techniques. We will further involve him/her in a client-based project which entails the application of these techniques to real-life problems.

We strongly believe that a close collaboration of all parties involved in DESCARTES will be of pivotal mutual benefit and will allow to form synergies that would be entirely inaccessible otherwise.

Sincerely yours,

Michael Spannowsky

Michael Spannowsky
AI4B LTD
8 Old Dryburn Way
DH1 3LE
Durham, UK

Phone: +44 7588 757 975
E-mail: michael@ai4b.co.uk

Date: 09.01.2020

Letter of commitment for DESCARTES
December 31, 2019

To: Prof. Konstantinos Vellidis, network coordinator
Institute of Accelerating Systems and Applications
University Campus, Physics Dept., Building IX, Zografou - Greece
P.O. Box 17214, Athens 10024, Greece

Letter of commitment

Dear Prof. Vellidis,

On behalf of Achilles Information Limited, I would like to express full support for the DESCARTES network programme, which aims at training early-stage researchers in the application of cutting-edge statistical learning models at the Large Hadron Collider data in the forthcoming years. We have carefully read the programme proposal and strongly subscribe to its contents.

We understand that you are interested in the development of ties between academia and industry to strengthen the training and research program of your network. Achilles Information Limited is enthusiastic to partner with the DESCARTES consortium and willing to collaborate in the following ways:

- We commit to host up to four trainees in our offices in Abingdon for periods of three to four months each.
- During the secondment, the trainees will receive training in cutting-edge machine learning techniques and close supervision from members of our organisation. In addition to the training and supervision, we commit to offer desk space, computing facilities and comfortable working conditions.
- We will train the early stage researchers in using Python programming for advanced machine learning tasks; we will teach techniques for reading vast volumes of data from relational, non-relational and graph databases. We will train the researchers on creating state-of-the-art visualisations.
- We commit to help the trainees with their career development plan; Achilles is very interested in considering the trainees as future hires in the Data Science team.
- We will participate in network activities by offering our expertise in statistical learning problems of interest for this project.

It is understood that the trainees will receive their salaries through the network.

We want to stress our interest in building strong ties with the academic partners in the network as we recognise the importance of training early stage researchers for the success of the High Energy Physics program at the LHC, and the role the vibrant field of data science and machine learning has in innovation and job growth in Europe. We thus wish you and the project success in this grant application and look forward to a fruitful collaboration.

Best regards,

Fabrizio Margaroli, Lead Data Scientist
Llanera, December 30, 2019

To: Prof. Konstantinos Vellidis, network coordinator
Institute of Accelerating Systems and Applications
University Campus, Physics Dept., Building IX, Zografou - Greece
P.O. Box 17214, Athens 10024, Greece
konstantinos.vellidis@cern.ch

Letter of Commitment

Dear Dr. Vellidis,

I would like to express our full support to the Advanced Statistical Algorithms for Particle Physics (DESCARTEs) project, which will provide a framework for training early-stage researchers on the advanced statistical learning and novel computing techniques needed for the analysis of large and complex multidimensional data samples, such as those acquired in the European Research Organization for particle physics (CERN) experiments.

At TWave we are interested in joining the DESCARTEs programme and collaborate with your network. We can offer the early-stage researchers our expertise in the technologies and tools related with our business activity, which is the development of highly integrated instrumentation equipment.

One of the work packages of your programme is aimed to the use of Machine Learning algorithms for full event reconstruction and imaging in particle physics and in industrial applications. This matches our R&D roadmap, as we are interested in ML applications on signal classification and fault detection.

We would be pleased to host two trainees from your network in our offices in Llanera, Asturias. This secondment will preferably take place in 2 periods, one trainee at a time. The evolution of the trainees’ research will be supervised by our personnel during the time they’ll be working in our office.

Best regards,

Juan Menéndez Blanco
Chief Technical Officer
Artabro Tech S.L.
Avenida de Santa María 73
15405 Ferrol (A Coruña)
Spain

Ferrol, 20 December, 2019

To: Prof. Konstantinos Vellidis, network coordinator
Institute of Accelerating Systems and Applications
University Campus, Physics Dept., Building IX, Zografou - Greece
P.O. Box 17214, Athens 10024, Greece
konstantinos.vellidis@cern.ch

Letter of Commitment

Dear Dr. Vellidis,

Artabro Tech is delighted to be part of the aDvancEd StatistiCal Algorithms for pARticIle physics (DESCARTEs) project, which will provide a framework for training early-stage researchers on the advanced statistical learning and novel computing techniques needed for the analysis of large and complex multidimensional data samples, such as those acquired in the European Research Organisation for particle physics (CERN) experiments.

At Artabro Tech we are interested in joining the DESCARTEs programme and collaborate with your network. We can offer the early-stage researchers our expertise in the technologies and tools related with our business activity, in particular those related to Artificial Intelligence.

Part of the DESCARTEs programme is aimed to the use of Machine Learning algorithms for full event reconstruction and imaging in particle physics and in industrial applications. The use of Machine Learning algorithms is at the core of our business. We will offer the trainee the possibility to apply advanced techniques outside academia.

We would be pleased to host one trainee from your network in our offices in Ferrol, Spain. The evolution of the trainee’s research will be supervised by our personnel during the time they’ll be working in our office.

Best regards,

Miguel Vidal Maroño
Chief Technology Officer
MARIE Skłodowska-CURIE ACTIONS

Innovative Training Networks (ITN)
Call: H2020-MSCA-ITN-2020

PART B

“DESCARTES”

This proposal is to be evaluated as:

[ETN]