**START PAGE**

MARIE Skłodowska-CURIE ACTIONS

**Innovative Training Networks (ITN)**

**Call:** **H2020-MSCA-ITN-2016**

PART B

“BigDAPHNE”

Big DAta in PHysics Network

**This proposal is to be evaluated as:**

 **[EJD]**

# LIST OF PARTICIPANT

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Consortium****Member** | **Legal Entity Short Name** | **Academic** | **Non-academic** | **Awards Doctoral Degrees** | **Country** | **Dept./****Division /****Laboratory** | **Scientist-in-Charge** | **Role of Partner Organisation[[1]](#footnote-2)** |
| Beneficiaries  |  |  |  |  |  |  |  |  |
| Aristotelio Panepistimio Thessalonikis | AUTH | x |  | x | GR | Dep. of Physics and IT Center Scientific Computing | Chariclia Petridou |  |
| Commissiariat a L'Energie Atomique | CEA | x |  |  | FR |  | Rosy Nikolaidou |  |
| University College London | UCL | x |  | x | UK | Dep. of Physics | Nikos Konstantinidis |  |
| Università di Pisa | UNIPI | x |  | x | IT | Dep. of Physics and Computer Science | Chiara Roda |  |
| Università del Salento | UNILE | x |  | x | IT | Dep. of Mathematics Physics | Stefania Spagnolo |  |
| Partner Organisations |  |  |  |  |  |  |  |  |
| European Organization for Nuclear Research | CERN | x |  |  | CH |  | Alberto Di Meglio | T,S |
| Consiglio Nazionale delle Ricerche | CNR | x |  |  | IT |  | Fosca Giannotti | T |
| ENEL Ingegneria e Ricerca S.P.A | ENEL |  | x |  | IT |  | Matteo Masotti | T,S |
| Georgia Institute of Technology | GTECH | x |  |  | USA | Dep. of Physics / Center for Relativistic Astrophysics | James Clark | M,S |
| GUARCO Interactive | GUARCO |  | X |  | IT |  | Guido Francioni | O |
| Heidelberg Institute of Technology | HITS | x |  |  | DE |  | Andreas Bauswein | M,S |
| Istituto Nazionale di Astrofisica | INAF | x |  |  | IT |  | Roberto Scaramella | T,S,M |
| Istituto Nazionale di Fisica Nucleare | INFN | x |  |  | IT |  | Giancarlo Cella | T,S,M |
| NCC GROUP | NCC |  | x |  | UK |  | Lloyd Brough | T,S |
| Semblent | Semblent |  | x |  | UK |  | Harry Howard | T,S |
| Software Carpentry  | Software Carpentry | x |  |  | US |  | Jonah Duckles | T |
| Transport for London | TfL |  | x |  | UK |  | Ben Cooper | T,S |
| THALES Service SAS | THALES |  | x |  | FR |  | Bernardo Resende | T,S |
| Université Paris Saclay | UNIPS | x |  | x | FR |  | Sylvie Pommier | T |

**Declarations**

|  |  |
| --- | --- |
| **Name (institution / individual)** | **Nature of inter-relationship** |
| C.Roda, V.Cavasinni - UNIPI | Research contract with INFN / user association CERN  |
| S.Spagnolo, A.Ventura - UNILE | Research contract with INFN / user association CERN |
| F.Fidecaro, I.Ferrante | Research contract with INFN  |
| C. Petridou, D. Sampsonidis, K. Kordas, S. Tzamarias - AUTH | user association CERN  |
| R. Nikolaidou, H. Bachacou, C.Guyot, J.F. Laporte - CEA | user association CERN  |
| Knowledge Discovery and Data Mining Laboratory (KDD Lab) – F.Giannotti, D.Pedreschi | Joint initiative of UNIPI IT Department and ISTI CNR |
| F. De Paolis, A. Nucita, F. Strafella - UNILE | Research contract with INFN |

**ABSTRACT**

We propose an innovative and interdisciplinary joint doctorate programme on fundamental physics with a twofold aim: (1) to explore some of the most fundamental questions in nature, building stronger links between the research fields of particle physics and cosmology; and (2) to train a new generation of savvy data scientists, who will benefit from the cross-fertilization between the academic and non-academic sectors in the programme to address common challenges in handling and analyzing large and complex datasets, and become leaders and innovators of Big Data Science solutions.

The research base of the programme are three experiments exploring from different angles questions such as the origin of Dark Matter and Dark Energy, and looking for signs of physics beyond the standard models of Particle Physics and Cosmology, using some of the world’s most powerful facilities: the Large Hadron Collider, a new generation of telescopes and state-of-the-art gravitational interferometers. The physics focus is on inter-related phenomena, while, on the technical side, all these experiments face the challenge of handling and analyzing vast datasets. The proposed ESR training will complement the standard physics PhD programme with a full suite of bespoke training activities, workshops and events, delivered by world-experts in the respective fields. By focusing the ESR projects on phenomena of common interest to the different research fields, the network will create a stimulating environment in which young researchers will broaden their scientific horizons, as well as acquiring a deep knowledge of their own research subject. Moreover, the training in data intensive science and technologies, reinforced with secondments in leading organizations in the field, will equip the ESRs with a unique skill set that is highly transferable and in high demand across both academia and industry, European and Global, in a diverse range of sectors.

#

# 1 Excellence

## 1.1 Quality, innovative aspects and credibility of the research programme

**Introduction:** The network research programme is based on three experiments exploring from different angles questions such as the origin of Dark Matter and Dark Energy, and looking for signs of physics beyond the standard models of Particle Physics and Cosmology, using the world’s most powerful instruments: the Large Hadron Collider, a new generation telescopes and state-of-the-art gravitational interferometers. The fast progress of the past 20 years brought us to a turning point in our understanding of the Universe: today we know that ~95% of the Universe is made of components of unknown nature, named - to partly reflect our ignorance of them- Dark Matter (DM) and Dark Energy (DE) - and accounting respectively for ~25% and ~70% of the content of the Universe. Understanding the nature of these phenomena requires new insight in the domain of microscopic constituents of matter and of fundamental forces and/or a revision of one or both of our most successful theories today, Quantum Mechanics and General Relativity. In addition, the era of gravitational wave astronomy has just started at the end of 2015 with the first observation of signals from the coalescence of black hole binary systems. A key prediction of the 100-year-old Einstein theory has been proven and a new frontier of experimental tests of strong field General Relativity is now open, bringing potential sensitivity to extended theories of Gravity. Large international research programmes have been established to overcome the limits of current investigation in all these fields: the LHC is pushing the high energy frontier in Particle Physics with an ongoing programme of data taking running until 2023 and a high intensity extension (HL-LHC) recently approved for the following decade; a number of space and ground-based telescope programmes underway and planned for the forthcoming years will map with unprecedented accuracy the early Universe; the second-generation laser interferometer for detecting gravitational waves, Advanced LIGO, is in operation since fall 2015, in the USA, and the corresponding European detector, Advanced VIRGO, will start operation before the end of 2016 near Pisa. **Based on these considerations, it is clear that fundamental physics stands at the beginning of a very interesting era.**

The most peculiar feature of the present research scenario is the absence of anomalies in data at accelerators that can suggest the optimal direction of future research, while a pressing need for an understanding of physics beyond the Standard Model of Particle Physics emerges from cosmological observations. At the same time Cosmology stands on foundations built with theories of Gravity and of fundamental interactions of matter. A successful exploitation of the rich panorama of experimental facilities available in the forthcoming decades and an optimal planning of next generation of experiments need the wide spreading of a solid shared competence in the scientific community working in the different experimental domains. Therefore, the Joint Doctorate programme **BigDAPHNE aims at building a new profile of scientist, tailored to the challenges of fundamental research topics, which sit at the intersection of particle physics, Cosmology and Gravity.** BigDAPHNE will offer first-class research opportunities to young researchers on high profile projects in fundamental physics, **ATLAS at LHC, Euclid and VIRGO**[[2]](#footnote-3), within the context of a common joint doctoral training program. The interplay between the different research domains will be a strong asset of the network, which will bring in a scientific cooperation forum, for the first time, researchers carrying out frontier research in each discipline across Europe. While the physics challenges focus on common and inter-related phenomena, on the technical side, all these experiments face the challenge of handling and analyzing huge datasets. Research in Particle Physics and Astronomy/Cosmology has been the biggest producer of large data volumes from around 50 years ago until recent times. *Today experiments in these fields produce data at a rate of hundreds of PB per year, comparable to the current yearly uploads to Facebook*. As a result, pioneering expertise on the methodologies for efficient data handling exists in these scientific communities. In the forthcoming years, frontier research programmes and new data taking campaigns will challenge the community with yet another significant increase in data volume requiring novel ideas for optimal exploitation. All the research projects within the BigDAPHNE network will face challenges related to large data sets and tiny signals to be extracted from an overwhelming background. Knowledge advancement will require a leap in the ability of researchers to effectively manage and analyse large volume of data. **BigDAPHNE will foster a strong cooperation on the methodological side between different research domains, focusing on advanced methods for data handling and data mining in science.** The researchers will share and exchange techniques which are so far domain-specific, they will use or develop novel methodologies taking inspiration from research outside physics and from experience outside academia.

A strong interplay between BigDAPHNE and the non academic sector can, actually, be implemented with great benefit on both sides. In the past two decades, the possibility to easily produce and collect information from a great variety of sources has tremendously increased. The scale of this phenomenon can be understood by observing that “the information created from the dawn of civilization through 2003 amounts to five exabytes while the same amount of information is now produced in two days”[[3]](#footnote-4). Already in the eighties, it was understood that the information hidden in this huge amount of data, nowadays referred to as “Big Data”, could be of extreme interest for the widest type of fields ranging from business to science, health care, security and policy. The ability to collect, store and analyse this ocean of data in order to extract useful insights has therefore become a very precious and rare skill. Indeed, **all observers agree that the abundance of data coexists with a profound shortage of data scientists**. *These professionals combine "the skills of computer scientist, statistician and narrator to extract the nuggets of gold hidden under mountains of data” and their jobs are proclaimed by The Economist to be "the sexiest job of the 21st century".* In particular, **Europe appears to be behind the USA, China, India and Brazil, in terms of the number of graduates with deep analytical training**[[4]](#footnote-5),[[5]](#footnote-6). The network will foster an active injection of expertise from the industrial and commercial side to the research side and viceversa, thanks to partnerships with companies leaders in Big Data technologies.

**In summary, BigDAPHNE can meet the goals of (a) increasing the potential for discovery in frontier research; and (b) populating the, so far, tiny community of EU data scientists, in response to the needs of both science and society.**

**Objectives:** The overall of BigDAPHNE is to build a new European Joint Doctorate (EJD) programme in Fundamental Physics and Big Data science to produce science professional with a high level of employability well trained to become leaders in academy and industry. The objectives of this innovative high-education model, listed hereafter, in-line with key recommendations[[6]](#footnote-7) to the European Commission and with a growing consensus in academia about the need for a coherent data strategy[[7]](#footnote-8) are:

1. To establish an **interdisciplinary** link between frontier research in Particle Physics, observational Cosmology and Gravity; the expected result is to train researchers in fundamental physics with an improved scientific profile: eclectic, highly receptive to new ideas and capable to identify and undertake the most promising research directions dictated by the implications of results from one field to another;
2. To foster the development of common advanced techniques to extract relevant features from large data sets; as a result, the researchers will be empowered by an enriched toolkit for discovery leading to faster progress in fundamental research;
3. To implement a pioneering model of research and training network where the strong exchange of knowledge and practices on Big Data methodologies between the academic and the non-academic sectors allows the next generation of scientists to undertake new paths to innovation;

**The format of a Joint Doctorate allows addressing at best the needs in terms of both training and cooperation in research that the objectives of the BigDAPHNE program entail**.

**Overview of the research programme:** Over the past decades, both Particle Physics and Observational Cosmology have experimentally proven the solidity of powerful models describing with high accuracy the current observations. In 2012, at the Large Hadron Collider (LHC), the long-searched for Higgs boson, a particle deeply connected with the theoretical foundations of the Standard Model (SM) of Particle Physics, has been discovered. No prediction of the SM has been contradicted so far, however this model describes only 5% of the content of the Universe and misses a Dark Matter candidate. Similarly, the Lambda Cold Dark Matter (LCDM) model provides a self-consistent description of the measured cosmological parameters, but the density of Dark Energy inferred from current observations is 1060 times larger than that expected from Particle Physics. This huge discrepancy cannot be reconciled without a fundamental re-evaluation of our understanding of the Universe. All Gravitational phenomena are described in the context of General Relativity and strongly speculative researches are trying to establish a common ground for the theory of Gravity and Quantum Field Theories. Extensions of Einstein’s Gravity can emerge as effective theories describing the low energy limit of models for the unification of fundamental interactions (like superstring theories, supergravity theories, GUTs) and they are also often invoked to solve some issues connected to the Mach principle, the dark energy problem, the coincidence problem.

The research projects of the ESR of the BigDAPHNE network will be developed in the contest of well established experimental programs, on the following subjects which exhibit a strong potential for connections between the domains of Particle Physics, Cosmology and Gravity:

**1) Scrutiny of Higgs boson properties and searches for dark matter and new phenomena** in proton-proton collisions at the center of mass energy of 13-14 TeV that will be delivered by the LHC. The data taking has started in 2015 and planned to last until the end of 2018. After the discovery of the Higgs Boson in 2012, a deep scrutiny of the opportunities for new advancements in Particle Physics at present and planned accelerator facilities was carried out, balancing scientific impact with costs and risks. The prominent role of the LHC in the roadmap for Particle Physics clearly emerged[[8]](#footnote-9). The LHC data offer the unique possibility to search for dark matter candidates and other new phenomena, and to perform precision tests of SM processes, especially connected to the gauge structure of the Standard Model and the Higgs physics. It is generally believed that the Higgs particle can play a crucial role in different scenarios of physics beyond the SM. As an example, the Higgs boson might be a “portal” connecting the known matter with a secluded sector of “dark particles” in scenarios of dark matter other than the WIMP (Weakly Interactive Massive Particle) paradigm in Super-Symmetry. Any deviation of the Higgs properties from the predictions of the SM might shade light on open questions about the origin of the mass, the matter-antimatter asymmetry, the role of the Higgs field in the universe inflationary era. Multiple production of gauge bosons will be scrutinised contributing both to the long standing campaign of precision measurements of the gauge couplings, started at LEP and progressing toward to goal of observing gauge boson scattering at HL-LHC, and to the search for phenomena like production of heavier Higgs bosons, of Gravitons, and of new gauge bosons from an extended gauge symmetry.

**2) Investigation of the nature of dark energy and dark matter** with the sky surveys from the Euclid space mission of the European Space Agency (ESA) planned for launch in 2020. Euclid was selected as the second medium-class mission in the Cosmic Vision programme in October 2011. The data collected by Euclid for the first time will lead to measurements of the cosmological parameters related to Dark Energy whose precision will not be limited by the statistical size of the survey. Therefore Euclid stands in a leading position in the investigation of the history of the expansion of the Universe[[9]](#footnote-10). It also plays a key role in the investigation of the nature and properties of dark matter and of the initial conditions seeding the formation of cosmic structures. The Euclid data will approach the precision required to start testing the unconventional scenarios where the signatures of the excess of matter and energy in the Universe are interpreted in terms of deviations of Gravity from the General Relativity picture. Two cosmological probes will be exploited and combined to address the scientific goals: Weak gravitational Lensing and Galaxy Clustering.

**3) Searches for signals of gravitational waves** in the data of Advanced VIRGO. The Advanced VIRGO experiment, operated by the European Gravitational Observatory (EGO), in Cascina (Pisa), is the most advanced laser-interferometer built in Europe for direct detection of gravitational waves. VIRGO and the U.S.Advanced LIGO experiments will operate jointly as a single large detector in order to maximise the reach of gravitational-wave astronomy in the next 5 years. The first detection of binary coalescence of black holes demonstrated that the threshold sensitivity for detection of these kind of sources has been reached and a relevant number of new BH-BH coalescence detections can be expected. This will allow testing models for the evolution of stellar populations and primordial black holes, with implications on dark matter[[10]](#footnote-11). Information about gravity in the strong interaction regime, including signatures of violations of General Relativity of cosmological relevance, and about the equation of state of neutron stars involved in a merger event can be recovered with a detailed study of gravitational waves from such events, that the Advanced VIRGO experiment is aiming to detect. Finally the study of non-Gaussian stochastic backgrounds of gravitational waves is a developing subject. It is based on analysis techniques common to the study of cosmologically motivated stochastic backgrounds, which can be challenged and commissioned already by the Advanced VIRGO data. It could put constraint on fundamental physics (inflation models and phase transitions in the initial stage of the universe evolution) though the timeline to reach the required sensitivity is the one of the third generation detectors on ground and of space detectors as LISA.

|  |
| --- |
| **Table 1.1: Timeline of the experiments in BigDAPHNE**  |
| **Year** | **Project** | **ATLAS** | **Euclid** | **VIRGO** |
| 2017 | Organization ESR recruitment | re-start Run2 | Analysis commissioning | Run |
| 2018 | ESR – PhD yr1 | End of Run2 with 150fb\*\*-1 | Analysis commissioning | Run |
| 2019 | ESR – PhD yr2 | LS1 – phase1 upgrade | Analysis commissioning | Run |
| 2020 | ESR – PhD yr3 | LS1 – phase1 upgrade | Launch |  |

The outcome of the 15 individual research programmes of the BigDAPHNE ESRs is expected to be the publication of new original results in the case of the ATLAS and VIRGO projects, and the delivery of a data-ready analysis strategy in the case of Euclid. The timeline of the experiments and of the ITN project is shown in Table 1.1.

**Research methodology and approach:** The activities in the project are structured in 5 work packages (Table 1.2). WP1 groups all the scientific research topics. WP2 has a mixed flavour: it will implement the structured training of young researchers on data science and will enhance the research programmes of the ESRs by reinforcing, updating and enriching the suite of data-handling and data mining tools for science. WP3, WP4 and WP5 are aiming at optimising the implementation and impact of the project and will be described in detail later in this document.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1.2: Work Package (WP) List** |  |  |  |  |
| **WP number and Title** | **Lead Beneficiary**  | **Start/End Month** | **Activity type** | **ESR** |
| 1 | Frontier Research in Particle Physics Cosmology and Gravitation | CEA | 8 | 43 | Research | 1-15 |
| 2 | Big Data Management and Analysis in Research and Industry | UCL | 8 | 43 | Research/Training | 1-15 |
| 3 | Training | UNILE | 7 | 48 | Training | 1-15 |
| 4 | Outreach and dissemination | AUTH | 8 | 43 | Dissemination/Communication | 1-15 |
| 5 | Management and coordination | UNIPI | 1 | 48 | Management | - |

*Addressing fundamental questions with the ATLAS, EUCLID and VIRGO research projects (WP1):*

The **ATLAS** projects in this WP will be based on analysis of the complete data-sample from the data taking that will last in 2017 and that will increase by a factor of about 20 the present data-sample. The peculiar feature of hadronic colliders is to allow searches in a wide range of final states. The proposed ESR research projects fully exploit this feature and thanks to the complementary competences of the various participating institutes we can propose searches in a wide range of final states The physics targets are: dark matter searches in final states with hadronic or leptonic activity and missing energy; Higgs properties and gauge coupling measurements or search for new resonances in fully leptonic or semi-leptonic final states produced from decays of pairs of gauge bosons; rare Higgs production modes and double Higgs production in finals states with b-quarks, t-quarks and τ leptons. Recently introduced algorithms for the reconstruction of boosted physics objects, like t-quark and W and Z, are often mandatory to achieve signal discrimination in the final states of interest to BigDAPHNE. While the official ATLAS software provides a standard reconstruction of all physics objects, the implementation of highly sensitive analysis strategies can require dedicated optimization of the object identification algorithms in addition to the use of powerful multi-input procedures for the statistical separation of interesting events from large background.

Research projects to develop techniques for dark matter and dark energy searches are proposed by Euclid. **Euclid** exploits the combination of redshift clustering and gravitational weak lensing by large-scale structures, typically referred as 3D weak lensing, to map the 3D Dark Matter structure of the Universe as well as its expansion history. Euclid will produce 5 PB of data for a 3 Billion-object catalogue, which will need extensive processing and will be complemented by large statistics simulations. The use of powerful statistical analysis methods is mandatory to exploit the full potential of these data. The Bayesian Hierarchical modelling and the use of new data driven statistics based on 3D spin-wavelets are innovative techniques, recently pioneered in preliminary studies[[11]](#footnote-12),[[12]](#footnote-13) and proposed for the interpretation of the EUCLID data in terms of Dark Energy. Extensions of the state-of-the-art compressed sensing GLIMPSE algorithm[[13]](#footnote-14) will be commissioned for Euclid Dark Matter physics with simulated data. In addition, the full power of research in cosmology comes from combining data from different sources: the development of automated pipelines[[14]](#footnote-15) for cross-correlating the weak and strong lensing signal with source catalogues from X-ray and gamma-ray space telescopes will be pursued to improve the combined physics reach in Dark Matter mapping.

**This part should be revised by Virgo’s collegues to introduce as primary objective the physic result**

In the context of **VIRGO** physics, the main tool for the detection and the study of merger is the Wiener filter technique. When unknown parameters must be determined, several template filters must be applied to the data. The number of templates must be large enough to approximate with the desired accuracy the manifold of admissible signals. The dimension of this template bank depends on the structure of this embedding, which can be quite complex especially when the effect of spin is taken into account. Several innovative techniques must be applied to reduce the complexity of the problem and the latency of the detection, as for example singular value decomposition, use of a wavelet basis and hierarchical approaches.

|  |
| --- |
| **Figure 1: Overlap of scientific and methodological activities** |
|  |

The complementary of expertise available within the network in terms of data analysis strategies, efficient modelling of multi-parameter effects (in signal and background processes at LHC, in gravitational wave damping and in cosmological observables) and statistical methods for model testing and marginalization will be exploited to provide solid guidance to each ESR’s project. The most profitable interactions among ESRs, as foreseen today, are described in Table 3.1d. Each beneficiary institution in the network will host researches acting in at least two different experiments of the network; therefore a strong and effective exchange of competences between different fields will happen not only in the fostered network interactions but also through the daily contact of ESRs enrolled in the BigDAPHNE project. With this interaction we aim at pushing the physics interpretation of the achieved results beyond the standard benchmark model testing, looking at implications and constraints from outside the specific domain of research by profiting of the broad-view approach to research fostered by the training in the joint doctoral program. Fig.1 schematically illustrates the research topics of each experiment highlighting the regions of clear overlap between them. There are domains (indicated by dotted and faint dotted lines) where several connections may be found, on a more more speculative ground.

In summary WP1 is designed to address objective 1 of the network project (see Section 1.1 Objectives).

*Catching the trends in Big Data strategies (WP2):* The work model in Particle Physics experiments at the frontier of energy and intensity, like ATLAS, critically depends on an efficient data strategy and heavily relies on distributed computing. Both reconstruction and physics result extraction have moved in the LHC era from simple cut-based techniques to machine learning (ML) techniques giving a significant improvement in the search sensitivities. Boosted decision trees, neural networks and other multivariate analysis (MVA) techniques are currently in wide use. However, the Particle Physics toolkit is still limited to a subset of the full suite currently on the market. The HL-LHC programme, where event complexity will increase by a factor of 10 and the data size collected for physics analysis will grow by a factor of 100 with respect to the current LHC data size and typifies the urgent case for widening the collection of Particle Physics methods and tools. At the same time, the development and commissioning of 3D spin-wavelet data compression methods and of Bayesian Hierarchical modelling, in the context of Euclid, and the determination of gravitational wave templates from extensive model dependent simulations of mergers are novel and complex data analysis methods highly demanding in terms of computing resources and data management strategies. The activities in this work package will aim at addressing objectives 2 and 3. They will develop on the solid ground of the innovative expertise in BigData technologies and methods established through the structured common training program (described in section 1.2) and will exploit strong partnerships with top level institutions in the academic domain, like CERN, INFN, INAF and CMCC, and companies operating in diverse areas, like Thales, NCC, TfL, Net7, COSMOTE’ where strong IT divisions manage large data analytics projects*,* and SAS, which is leader in the deployment of integrated software solutions for Data Mining widely used by commercial and industrial clients.

The Information Technology division of CERN played a leading role in developing the IT strategies and technologies for the research in Particle Physics, in cooperation with world leading companies and other research centres, including INFN. In recent years, a strong synergy with the non-academic ICT research has been established through the OPEN-Lab consortium, where R&D programmes of common interest for CERN and world leading ICT companies are pursued in full co-operation by researchers employed by CERN and by partner companies like INTEL, ORACLE, Rackspace and others. In the nineties, when the Internet started to offer significant bandwidth, the cloud-computing model emerged from non-academic development area. The strong increase in volume and traffic of data from human activity footprints on the net, overcoming the volume of scientific data, built a case for the implementation of data cloud services from commercial providers: Amazon since 2002 and later Google since 2009. Similarly, a wide range of powerful data-mining techniques has emerged from diverse sources outside the pure research environment or in IT research activities decoupled from the research efforts in physics. The INDIGO-DataCloud project, funded by H2020 (grant agreement RIA 653549), lead by INFN, is developing a new user-driven open source data and computing platform targeted at various scientific communities.  Among the services it provides, the “Data Mining and Analytics for eScience” tools are based on Ophidia, a framework for data intensive analysis, developed within the context of CMCC activities and becoming of wide use in various scientific domains, from Astronomy to Environmental and Hearth Science up to Social Science. *A dedicated training on the Ophidia toolkit, as well as on the SAS Analytics framework, will be concrete actions addressing the observation that the scientific communities working on sky surveys, in experimental Particle Physics and Gravity share a surprisingly limited set of techniques in the practice of research.* A recent workshop[[15]](#footnote-16) held at CERN brought to a non-expert audience the interesting opportunities that might arise from a deeper communication between different forums. **Work package 2 of BigDAPHNE will contribute to closing the gap between technical expertise on data treatment in different areas, including non-research applications**. Its main objective is to inject in the training of young researcher’s specific deep competences on data handling technologies and data analysis. **ESR will be fostered to** **apply in their research programme novel data analysis techniques borrowed by other applicative or research fields aiming at standardising valuable tools in areas where their use is not yet common**.

The ESRs of BigDAPHNE will consider for application in their individual research projects:

1. Efficient analysis techniques and optimised data access methods which are of wide use in other research domains within the network (Particle Physics researchers can evaluate wavelet analysis or automated pipelines, VIRGO ESRs can consider MVA techniques in use in Particle Physics);
2. Modern machine learning techniques (i.e. unsupervised learning techniques, deep machine learning, vision classification algorithms, etc.) well integrated in the suite of tools in use by research domains other than physics or outside academia, taking advantage from the expertise of academic (CERN, CMCC, AUTH grid data-center) and non-academic (SAS, ENEL, Thales, NCC, TfL, Net7) partners;
3. Analysis models inspired by Big Data computing frameworks (Hadoop, Spark, Cassandra, etc.) for a fast turnaround of the analysis loop and a reduction of the typically big time-investment in analysis framework definitions (ESRs working in ATLAS might elect a single, or a few, analysis framework(s) addressing their general needs while optimised for scalability, maintainability and easiness of integration onto new or existing analysis toolkits);
4. Innovative techniques already emerging in their respective research environment (examples have been mentioned for the Euclid and VIRGO experimets);
5. Commonalities of the computing approach in domains with different data models (Particle Physics and sky surveys) to identify opportunities for synergies in developments of common interest.

A potential breakthrough would be the foundation of an open source library of algorithms implemented in a non-discipline specific manner for wide use in science and outside.

### **Originality and innovative aspects of the research programme:** The academic community in Europe is historically strongly involved in leading research in Particle Physics, in Observational Cosmology and in Gravity. However while the results of the last couple of decades has brought the research topic to increasingly overlap, a wide view of these research domains remains a prerogative of a few selected scientists and EU still lacks of a project to train scientists with this skill. Moreover top-level expertise in the traditional analysis methods involved with these kinds of researches is wide spread. However, while detector design, to make an example, is widely felt as a research topic and therefore included in research and training paths, the expertise in computing, software and data analytics design is typically treated as an accessory skill, to be gained by the researcher through (sometimes bad) practice. The fast growth of the experiment data size and complexity makes computing and data manipulation methodologies fundamental research tools, deserving on their own intellectual investment for development and optimisation. Recently, some innovative research paths have been traced at the border of physics research in the EC-funded ICE-DIP[[16]](#footnote-17) EID and the HPC-Leap[[17]](#footnote-18) EJD programmes. The focus of these projects is on high performance computing for scientific research. The newly established AMVA4NewPhysics[[18]](#footnote-19) ITN instead aims at developing thematic advanced statistical tools for specific use at LHC.

### In the present panorama, BigDAPHNE stands as a distinct interdisciplinary Joint Doctorate programme in Fundamental Physics and Data Science whose main point of innovation are:

### **1. pioneer of a model of high-education network where scientists from highly competitive experiments in different domains, all investigating the central topic of the structure and content of the Universe, join in a training and research environment aiming at preparing a wide-open vision of research in fundamental physics for the next generation of researchers;**

### **2. build a common ground of innovative and cross-disciplinary methodologies for research in the area of data analytics, exploring the trends and opportunities emerging from outside physics research and from the non-academic sector.**

## 1.2 Quality and innovative aspects of the training programme

### **Overview and content structure of the doctoral programme:** The training programme provided by the network aims at forming a new profile of scientist, with a wider knowledge on the compelling open questions in fundamental physics, expert in cutting-edge tools on data science, able to export his/her technical expertise in data science to different disciplines and contexts and capable to lead projects in an international multi-disciplinary environment. In order to reach these goals, the training focuses on three pillars:

**1.** **PhD research projects** on pressing open questions about particle physics, cosmology and gravitation in international collaborations with world-experts from two institutions in two different countries. The network will require fellows to interact, exchange results and ideas pushing them to find links and complementarities among the different research environments. While the results of their individual research project will be described in their thesis the ESRs will report the result of this collaborative work in a document describing the connections among the research projects of the network and ideas on how to build a common research ground (Mxx).

**2.** Structured courses, short workshops and seminars to build candidates with strong profiles non only on **fundamental physics** by also on **good software practices, Big Data technologies, statistics, business mind-set for data projects (ideas for data projects, experience from members of the companies, etc.) and communication (presentation, writing a cv, writing a project, etc.)**. The aim of widening the fellow’s physics background will be fostered through a cycle of **online seminars** given by world-experts in the various disciplines. The BigData courses will provide skills to help fellows to be both more efficient in the research work and better formed for data scientist positions in the industrial or academic environment.

**3.** **Secondments (3-6** **month long) at academic or non-academic partner institutions leaders in Big Data technologies or power users of cutting-edge data analysis strategies**. Each ESR will be hosted in one of qualified partner institution to work on a BigData project under a qualified supervision. ESR will contribute to business analytics data-projects in various fields or will develop tools in the R&D activities. The non-academic stakeholders hosting ESRs are world-leading institutions active in the most varied sectors of Big Data (mobility, internet security, Environment Science, telecommunications…). The academic institutions hosting secondments (CERN and AUTH-IT center) are leaders in IT technologies and Big Data managing methods and they have close connections with the industry. A preliminary list of the secondment projects proposed by non-academic and academic institutions is given in Table 1.2c.

**The training by research in a cross-disciplinary environment, the dedicated and structured training in data handling and data mining methodologies and the secondment experience are all together innovative aspects that add a valuable component to ESR’s curricula that is missing in standard PhD training.** The structure of the training activities in the BigDAPHNE programme will provide a valuable background in the new born discipline of Data Science and a practical experience in the application of this knowledge to the non-academic world. In addition, a deep understanding of the thesis research field and the ability to manage independently a research project will be acquired. On the successful completion of the programme the ESRs will be granted a double degree.

### All the ESRs will be recruited at the same time. We assume to be able to have all ESRs recruited by M8 and we aim at enrolling them at the 2018 doctoral programmes (Table1.2a). The ESRs will therefore be able to participate to the network-wide training courses at the same moment. Special ESR requests to participate to additional courses or seminars will be considered and accepted if judged suitable/feasible.

**Table 1.2 a - Recruitment Deliverables per Beneficiary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Researcher No.** | **Recruiting Institution / Institution A** | **Institution B** | **Planned Start Month** | **Duration**  |
| **1.** | **UNIPI** | **UCL** | **8** | **36** |
| **2.** | **UNIPI** | **CEA** | **8** | **36** |
| **3.** | **AUTH** | **CEA** | **8** | **36** |
| **4.** | **AUTH** | **UNILE** | **8** | **36** |
| **5.** | **UCL** | **UNIPI** | **8** | **36** |
| **6.** | **UCL** | **UNILE** | **8** | **36** |
| **7.** | **CEA** | **AUTH** | **8** | **36** |
| **8.** | **CEA** | **UCL** | **8** | **36** |
| **9.** | **UNILE** | **AUTH** | **8** | **36** |
| **10.** | **UNIPI** | **AUTH** | **8** | **36** |
| **11.** | **AUTH** | **UNIPI** | **8** | **36** |
| **12.** | **UCL** | **UNILE** | **8** | **36** |
| **13.** | **CEA** | **UCL** | **8** | **36** |
| **14.** | **UCL** | **CEA** | **8** | **36** |
| **15.** | **UNILE** | **CEA** | **8** | **36** |

### **Network wide training events:** The events organised by the network are cross-disciplinary and compulsory for all ESRs. The hosting institutes will take care of the local organisation. A programme committee will be setup for each event that will always include the WP3 leader. This committee, in coordination with the Academic Board will be responsible for the event programme. The event material will be published on the BigDAPHNE site. The training events will be also open to fellows external to the network up to the maximum capability. The quality of the training of each event will be assessed with a questionnaire and feedback from these assessments will be sent from WP5 leader to the Academic Board (described in section 3.2).

The network wide training events are listed in Table 1.2c and cover both training in fundamental physics and BigData domains. **Event 1** is a cycle of about 16 online seminars, with a bi-monthly frequency, given by world expert professors, selected in the international community, on the foundations grounds of the research subjects addressed in the ESR projects. These seminars are complementary to the courses organized by the PhD schools and will form a common background on gravity, cosmology and particle physics to the ESRs. The online infrastructure will be provided by e-learning framework developed by the members of the Engineer Innovation Department of UNILE[[19]](#footnote-20). **Events 2,3,4** are sequential schools which aim at growing the expertise in BigData technologies. The schools are designed to profit both from expertise within the research environment (INFN, CERN, CMCC) and from a company leader in the data science sector (SAS). The first school (**Event 2**) aims at providing the students with foundations on good programming practices, computer architecture and their use, parallel computing and high performance computing. Selected seminars on BigData aspects from various research and applied environments will complete this first school. INFN IT experts will organize the school that will be based on the Bertinoro INFN school[[20]](#footnote-21) program appositely enlarged and modified to satisfy the need of the network. **Event 3** is a school with mixed flavour and will provide a more complete view on Big Data technologies together with introduction to Data Mining for market applications from SAS (5 days course) and for research environment with tools developed within INDIGO-DataCloud EU project in collaboration with CMCC institution. This event will be organized by SAS members and by IT experts from UNILE and CMCC. **Event 4** aims at giving a complete and exhaustive training on BigData technologies in particle physics, cosmology and gravitation astronomy. The organization of this school is leaded by CERN members, which have a huge experience in IT school organizations. The UCL, INAF, CEA and AUTH members will provide the additional expertise on technologies and methods used in cosmology and gravitational astronomy. Events 2,3,4 are central training activity for the ESRs and we plan to have ECTS credits assigned to each of them requiring final assessment exams. We aim also at preparing ESRs fellows to take the exam for SAS Data Mining certification.

**Table 1.2 c - Main Network-Wide Training Events, Conferences and Contribution of Beneficiaries**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Main Training Events & Conferences** | **ECTS** **(*if any*)** | **Lead Institution** | **Project Month (*estimated*)** |
| 1 | Cycle of online seminars – Highlights and foundations on particle physics, cosmology and gravitational astronomy |  | UNILE | M10+2\*n, n=0,…15 |
| 2 | First Network school: “Foundations – effective programming, high performance computing and Big Data highlights” | 10 | INFN | M9 |
| 3 | Second Network school: “Introduction to Data Mining techniques in society and research” | 5 + SAS DataMining certification | SAS/UNILE/CMCC | M17 |
| 4 | Third Network school: “Data science in fundamental physics” | 10 | CERN/UCL | M30 |
| 5 | 2 days workshop “An introduction to Hadoop” |  | AUTH | M30 |
| 6 | Workshop on “Big Data: a wide view from industrial and academic projects” |  | AUTH | M30 |
| 7 | Conference on scientific harvest |  | UNILE | M41 |

The EJD fellow’s training will be completed with two-day workshop on Hadoop (**Event 5**) held in Thessaloniki and organized by AUTH-Grid. Expertise with this framework is a must for the profile currently requested on the market of Data Science jobs. Members of the AUTH grid and computing centre are experts in Hadoop and have already delivered courses on this framework. The AUTH-Grid centre will provide the necessary infrastructure for hands-on laboratories. Event 6 and 7 are dissemination events and are described in Section 2.3.

During the annual network meetings or during the network schools, training devoted to **transferable skills** will be organised. This will include training in presentation skills, project and scientific writing, CV writing. Many institutions in the network have expertise on organisation and delivery of these type of courses (INFN, CEA, UCL, etc.) moreover we will take advantage of the presence of two Royal Society fellows to profit of courses organised by the Royal Society. Seminars on building a business mind for data project will be invited at each major event of the network (annual meetings, main schools). ESR will consolidate communication and managerial skills by defending their work in large and competitive groups, in front of an expert audience, at conferences, or addressing the general public in outreach activities, or transferring knowledge to other fellows or students through the network research and training activities.

**Role of non-academic sector in the training programme:** The strong involvement of members from the non-academic partners to the **organisation, supervision and training** of the fellows is one of the distinctive features of this network. The industrial partners of the network offering secondments are either word-leading developer of Big Data tools or users of cutting edge technologies to extract profit valuable information from their data in the most varied fields: mobility, telecommunication, internet security, etc. Industrial partners will provide expertise and a different view with respect to the academic one. In addition to hosting ESRs, the non-academic partners will contribute to the network with seminars meant to give a wide view of the type of projects and skills that are interesting for companies and to explain how to develop a business data project. The participation in the training from a leading company in Data Science (SAS) is also a very innovative and important characteristic of the network. They will inject in the training up-to-date trends and requests from the market in addition to tools and visions complementary to those developed in the research environment. In order to foster the ESR’s entrepreneurship, we will profit in particular of the experience of people that from an academic career moved to business, starting a new company or joining a big enterprise. Members of the network partners and other non-academic stakeholders will play a key role in these activities.

**Table 1.2c – List of secondment data projects – TO BE UPDATED**

|  |  |  |
| --- | --- | --- |
| **Stakeholder hosting secondment / Country** | **Subject** | **Gained Expertise** |
| ENEL Energia e Ricerca SPA (ENEL) / IT | Energy Management for trading department | data mining, …  |
| Smart City Malaga – Consumption profile |  |
| NCC Group NPC (NCC) / UK | Network activity data management and online analysis  |  |
| Advanced network activity analysis |  |
| Thales Service SAS (Thales) / FR | Tools for Big Data services |  |
| Securing IoT: data protection in new landscape |  |
| Transport for London (TfL) / UK | Predicting and preventing train failures |  |
| Optimization of London Underground asset maintenance strategy |  |
| Information Technology Center of AUTH (AUTH-IT) / GR | Intrusion detection and alerting service  |  |
| Environmental monitoring and energy efficiency |  |
| Semblent / UK | Auto query generation with Gremlin  |  |
| Efficient export of graphs to relational data |  |
| European Centre for Nuclear Research (CERN) / SW | Big Data analytics techniques for High Energy Physics analysis |  |
| Tools to support unsupervised machine learning  |  |
| Real time technologies for large scale analysis using Cloud deployments |  |

## 1.3 Quality of the supervision

**Qualifications and supervision experience of supervisors:** ESRs will be supervised for their research work by two supervisors, one from each university where they will obtain the PhD. In some cases, mentors from our academic partners will provide additional research expertise to complement the PhD supervision. Supervisors and mentors have been chosen to ensure full and highly competent coverage of all aspects of the research programme. All the supervisors have long experience in supervision of PhD research projects and have leading positions at international level in the research topics they will supervise, ensuring both the necessary competence in the research subject and good connections that will allow guiding the ESRs in the academic and research community. All ESRs will also have a mentor from the non-academic or academic partners where they will be seconded. Besides adding a vision of the Big Data aspects in the commercial world they will offer technical help during all PhD and guidance during the secondment period. The supervisors and mentors assigned to each ESR projects are listed in Table 3.1d. A short profile of all supervisors and mentors listing their experience in supervision and expertise in research or technical aspects can be found in Table 5.

### **Quality of the joint supervision arrangements** At the beginning of the project ESRs together with the mentors and supervisors (Supervisory Team described in Section 3.2) will compile a Career Development Plan (CDP) to define scientific and training objectives and the timeline for their achievement. Meeting with the supervisors will be frequent and at minimum once per week to allow guidance in the research and monitoring of the progress. These meetings will also allow coordination among the supervisors and mentors. Official monitoring and assessment of the ESR progresses will be done quarterly by the Supervisory Team through a presentation by the ESR with slides and through a written report annually at the meeting with ST including the AB chair (see Table 3.2b). Slides and reports will be collected on the web site to document the progress. Annually the ST together with AB chair will produce a report assessing the advancement based on the CDP. ESRs will gain a double degree from two institutions (“Institution A and “Institution B”) as listed in Table 1.2b. The students will split their research time about equally between the two institutions of the double-degree.

## 1.4 Quality of the proposed interaction between the participating organisations

**Contribution of all participants to the research and training programme**: All nodes, including the partner organisations, will participate in the PhD training programme. Most institutions participate with multiple actions; as an example, industrial partners not only offer secondments, but also provide training on non-academic aspects on BigData or career developments outside academia. UNIPI, UCL, AUTH, Université Paris-Saclay[[21]](#footnote-22) and UNILE provide the PhD courses and infrastructure to deliver the double PhD degrees and offer strong and complementary expertise relevant for the ESR scientific projects. Members of these institutions cover leading roles in various research areas within ATLAS, EUCLID and VIRGO as described in Section 3.2. The universities collaborate on research with important research institutions (INFN, OAR of INAF, HITS…) that offer not only additional expertise on research aspects with contributions to supervision (see Section 3.2) but also use of their cutting edge research infrastructure.

The crucial role and contributions to the training activities of SAS, CERN, AUTH-IT, CMCC has been discussed in Section 1.2. The role of the industrial partners in the training of the ESRs is equally vital. Companies from different countries (Thales, Net7 , OTE/COSMOTE`, SAP, NCC, TfL) with strong IT divisions managing large data analytics projects in the area of transport, telecommunication, security, geographical analysis will assign projects to ESRs injecting in the network a vision of data analytics from a perspective alternative to research. In addition, members of the companies will mentor the ESRs on career aspects from a non-academic perspective, they will also offer insight in the use of Big Data in a commercial environment, and will provide training at joint workshops on the tools and techniques used in industry for data mining, data managing, data presentation.

**Synergies between participants:** The consortium pools competencies to a higher level of excellence than what is available at the individual nodes for the benefit of the training of the ESRs and of the research environment. The network builds on long-standing research collaboration between members of the different institutions within the experiments (CEA, UNILE, AUTH have been sharing leading responsibility and roles in muon identification and triggering in ATLAS; AUTH, UNIPI, UCL have a strong collaboration on track based triggering in ATLAS, UCL and CEA share leading responsibility in the Euclid consortium; CEA lead an ITN network [funded by … ] on Higgs Physics where UNIPI, AUTH and UCL were participating) and, on the other hand, puts a large effort to enlarge and enrich the existing collaborations with the aim of:

* building a new long-standing double-degree PhD path;
* growing a new research community pulling together complementary research fields;
* creating new two-direction connections with companies for training and research activities.

The four-year ITN project will be the seed for longer-term collaborations and for future more challenging projects.

**Exposure of recruited researchers to different (research) environments, and complementarity thereof:** The network contains a variety of institutes of various sizes that the ESRs will be exposed to. The ESRs will experience universities, research institutes, international collaborations and industry workplaces, covering the full range of environments they can expect to encounter in their future careers. They will spend time at two world-leading universities, in two different countries, as part of their joint doctorate degree, which will provide invaluable experience of research training in different European countries. Additionally, each ESR will carry out an internship at a partner company, immersed in one of a wide range of industrial work places: finance, cyber-security, telecommunication, transport and consulting. The interaction between the ESRs in various events of the network, and particularly in the 3rd annual meeting, where they will each present their experiences from their internship projects, will enhance their learning experience from the non-academic environment and better prepare them for their future careers as leading data scientists.

# 2 Impact

The impact of BigDAPHNE will be twofold: on **fundamental research** and on **training**.

On **fundamental research**: BigDAPHNE will address the key questions about our universe on Dark Matter (DM) and Dark Energy (DE) in a coordinated, coherent way across the fields of particle physics (PP) and cosmology (Co), by

* Exploiting and developing further the synergies between the two fields and by bringing together renowned scientists in these areas and coherently educating and training competitive and promising young researchers towards their PhDs, in both fields.
* Involving and developing further cutting edge technologies and methods, in the sense of bringing together the most modern and important experiments in these fields, and exploiting their common denominator, the Big Data, from the handling and analysing the data, to the extraction of the results.

**BigDAPHNE will become an example of close collaboration between PP and Co and is expected to provide a break through in addressing the key questions on DM and DE by opening new paths in this research and by triggering and contributing to discoveries.**

On **training**: The training of BigDAPHNE is concentrated on Big Data. Big Data are nowadays produced in almost every field of science as well as in business, economy and fundamental research. To extract useful results from the bulk of those data special capability is required, which has to be urgently developed. **Big** **Data handling therefore is an emergency,** which has to be rapidly evolved to cover the needs of the academic and non-academic world.

**The project will offer a modern training programme that will keep pace with the rapidly evolving needs related to Big Data in science and society; will produce the human capital to impact the competitiveness of Europe in science and will enhance the ability of European industry to pioneer development**.

Furthermore, BigDAPHNE, by creating the necessary infrastructure between the collaborating institutes, will pioneer in establishing a modern doctoral degree programme with novel training, involving both inter- and multi-disciplinary fields. The project will take advantage of the bright minds attracted more and more in the fields of PP and Co because of the recent advancements in these fields, and will try to **sustain this modern approach to doctoral degrees in Europe,** **beyond the funding period of the project.** Finally, the consortium, acknowledging the impact of visual/video methods in dissemination, will introduce a documentary film, among other outreach activities, as the utmost procedure to **communicate challenging scientific environments and novel training** **to the general public and mainly to high school and university students.**

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

**Impact on academic career:**  The ability to use and implement new Big Data techniques is becoming imperative to the research groups in the fields of particle physics, observational cosmology and gravitation. The ESRs enrolled in the BigDAPHNE programme will acquire high-level training on Big Data, which they will apply to their respective fields of research with the possibility to make breakthroughs in the scientific knowledge, thus becoming strong candidates for academic jobs with key expertise that is not yet widespread in the research domains.

**Impact on non-academic career:** The exposure in Big Data managing and Big Data analysis tools will provide cutting-edge knowhow and will open entirely new perspectives to the career of the ESRs.

The strong involvement of the non-academic sector in the training programme and the dedicated secondments at the companies’ premises will create a pool of high-level researchers, with unique expertise, highly in demand by the European Companies. The formed ESRs, as new PhDs, can flourish in the non-academic sector as developers in data mining, as Big Data analysts, in business and market research to give only a few examples.

**Impact on entrepreneurial skills:** **The exposure of the researchers to the non-academic environment through their secondments, will allow them to discover opportunities to convert knowledge and ideas into products and services for economic and social benefit and to develop entrepreneurial skills.** Specialised training events are foreseen to actively involve the ESRs as main actors on the scene of enterprise.

**Impact on mobility:** The mobility of the researchers is a built-in component of this project. The BigDAPHNE network involves several EU countries increasing thus the internationalisation of the participating organizations. The researchers will be exposed to research and industrial practices that do not belong to the same country thus acquiring a multi-cultural work experience which will make them able to operate across borders and beyond cultural barriers. This exposure will also challenge them to initiate new practices at home, contributing to the spread of development and innovation throughout Europe and to the harmonisation of educational and research practices between countries.

## 2.2 Contribution to structuring doctoral/early-stage research training at the European level and to strengthening European innovation capacity, including the potential for:

### **Contribution of the non-academic sector to the doctoral / research training:** The BigDAPHNE joint doctorate programme provides a great opportunity to build the necessary bridge between academia and industrial partners, at the level of high education, which will be beneficial to both. The non-academic sector, when looking for highly qualified employees, often suffers from the too abstract training provided by standard education paths. On the other hand, traditional PhD schools do not always guarantee strong opportunities for employment to their trainees. The integration of industrial partners in the training and research programme represents a priceless ingredient with respect to the standard high education schools in Europe. **The ESRs trained in BigDAPHNE will gain experience both in fundamental research and in data project applications during secondments, acquiring expertise in a domain with a high growth potential in the private sector.**

### **Contribution to developing sustainable joint doctoral degree structures:** Although joint doctoral degrees are quite prestigious a major obstacle for their implementation is the difficulty to harmonise rules among different doctoral schools. BigDAPHNE will create the necessary infrastructure between the collaborating institutes and will continue the double degree programme beyond the funding period of the project. Furthermore, will make an effort to sustain the funding beyond the duration of the project. While the academic sector will seek funding from the existing mechanisms national-wise, the involved industrial sector will open new potential funding streams to complement this, thanks to their strong research-intensive interest on Big Data. Overall, the experience gained with BigDAPHNE will be an excellent basis for similar cases in the future.

For example, other disciplines in the same educational units will profit from the creation of the legal framework and use it as well. Such structures can attract talented students worldwide and give them the opportunity to be exposed to competitive multicultural research environments.

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

### **Dissemination of the research results**: BigDAPHNE will follow the standard international practice for the results of the scientific research that is to publish them in high impact scientific journals and to present them to International Conferences and workshops of the respective fields. The cross-disciplinary nature of the consortium will enable dissemination of the Particle Physics results to the Cosmology community and vice-versa. Results on new software tools and techniques on data handling will be published in computational journals. The network will play also an active role in organising two events (event 4 and 7 in Table1.b) dedicated to dissemination of technical and scientific results to a wide community:

* **Big Data: a wide view from industrial and academic projects**: We will hold a [**joint industry/academia workshop**](http://depts.washington.edu/dswkshp/) where Big Data projects from different communities will be discussed. In this event we will advertise the project and its benefits to a wider audience with the aim of fostering more inter-sector and inter-field co-operation. Articles will also be uploaded to platforms such as International Science Grid, which has a mixed audience of industry and academics from a range of fields.
* **Conference on scientific harvest**: We will hold a final conference on the scientific aspects of the network open to a wide international community to foster the link between particle physics and cosmology.

### All public results, both on fundamental research and on new Big Data tools and analysis, will be available on the dedicated website of the project. In line with the EC **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**, the consortium will use open access archives and journals for publication, like [arXiv](http://arxiv.org) and [INSPIRE](http://inspirehep.net).

### **Exploitation of results and intellectual property:** There is no doubt that the scientific results delivered by the young scientists within their research teams will lead to further outstanding research activities, both for the teams involved and the related experiments. Furthermore, having the industrial involvement in the doctorate programme well intertwined with the research projects, will ensure the consolidation of scientific knowledge relevant to applications and innovation. Companies like SAS, SAP, CMCC and the International research center CERN, through the trainings and secondments they will provide, will both gain in their respective research areas and also motivate researchers outside academia to the high impact Big Data challenge. Companies like TfL, NCC, Thales, COSMOTE, Net7, can further exploit software developments related to Big Data handling and Big Data analysis that will derive from the project. Already, through the secondments of ESRs, new commercial products or improved services to the public may become available. The involvement of industrial partners requires a fair policy on intellectual property rights (IPR). The Consortium Agreement of the project will handle IPR issues. The rules will encourage sharing of the foreground in the consortium, and at the same time will protect the participants to be able to exploit resulting intellectual property.

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

### **Communication and public engagement strategy of the project:** The ESRs will play the leading role in communicating the project to various key audiences: schools, undergraduates, industry, policy makers, academics and general members of the public. The strategy will include both targeted and more general means of communicating the project’s aims and successes. The more general measures will include:

1. The project website will host blogs regularly updated by the ESRs, short interviews and ‘success storyboard’ to emphasise research highlights and equal opportunities in research.
2. A documentary film will be made to explain in a simple and pedagogical way the main points of the research done in particle physics and cosmology, what Big Data is and its direct application in every day life. The goals of the documentary will be a) to teach the young generation the challenge, importance and excitement in fundamental research and the progress that can be made working closely with the non-academic sector and b) to educate the public how fruitful the collaboration between academia and private sector can be, to the benefit of science and society. Such a documentary will afterwards be distributed to larger audience, will be uploaded to the web, as well as will be shown in documentary film festivals (e.g Thessaloniki documentary firlm festival).
3. The project will be promoted at science exhibitions, such as the [Royal Society Summer Science Exhibition](http://sse.royalsociety.org/2015).
4. ESRs will promote their research at events such as the [European Reaserchers’ Night](http://ec.europa.eu/research/researchersnight/index_en.htm) (ERC) and [Science Museum Lates](http://www.sciencemuseum.org.uk/visitmuseum/plan_your_visit/lates.aspx). They will also be encouraged to attend events such as [Pints of Science](https://pintofscience.com/about/), Science Caffe` which aim to target a wider audience who may not usually engage in science outreach activities.
5. Articles will be targeted at both the mainstream (for example: [The Guardian](http://www.theguardian.com/science/life-and-physics) or CERN-Courier) and the online press.

Events will also be undertaken at the nodes to target the aforementioned key audiences:

1. Schools: All the ESRs will be expected to present their research at a minimum of one school a year. A Big Data app will be produced to provide a hands-on demonstration of science. The ESRs will also participate in [MasterClass](http://physicsmasterclasses.org/) events at their nodes.
2. Undergraduates: [Meet a researcher](https://www.ucl.ac.uk/brain-sciences/student-new) style events will be hosted at the nodes to encourage more students to undertake research careers. These events will also focus on how the ITN helps improve their training and employability.
3. Policy makers: Attend [events](http://www.bivda.co.uk/Blog/tabid/1520/entryid/70/parliamentary-links-day-2015.aspx) focused on promoting science and its benefits to local associations and elected representatives, such as the [Parliamentary Links](http://www.bivda.co.uk/Blog/tabid/1520/entryid/) events in the UK.

# 3 Quality and Efficiency of the Implementation

## 3.1 Coherence and effectiveness of the work plan

The work plan is organized in two research/training WPs plus the three organisational WPs for training, dissemination/outreach and management (table 3.1). The first WP focuses on research in **“Particle Physics,** **Cosmology and Gravitation”**. The proposed ESR research projects are designed to address some of the most pressing questions in these key sectors of fundamental science. They are inter-related and in most cases they approach the same problem, e.g. Dark Matter, from different angles, a strategy that both has the potential for higher impact and gives the opportunity to the ESRs to broaden their scientific horizons beyond their specific research domain and benefit from breaking the barriers between research domains that are commonly pursued in isolation. The effectiveness of the WP1 implementation is ensured both by the strong, complementary expertise offered by the ensemble of beneficiaries and partners of the network and by the diverse expertise and experience in project management of the WP lead beneficiary (CEA). Section 5 describes in detail the world-leading expertise that we have gathered in the network.

The WP2 has a mixed research and training flavour and deals with the technical challenge of handling and analysing vast datasets, a challenge that is common to all the ESR research projects, as well as to the activities of all the industrial partners of the network. The aim of this WP is to harvest the synergies between the different research domains and between academia and industry in Big Data Science and technologies, to transfer and apply novel data science techniques from one field to another, and to train the young researchers in the network to become leaders and innovators with vehicle these tools and techniques. The separation of (the technical) WP2 from (the research-focused) WP1 will give the correct emphasis to the interdisciplinary nature of the data handling and analysis methodologies. The ESRs will profit from world-leading experts in these technologies, both from the academic side (CERN, CNR, IT Department AUTH), as well as from the industry world (SAS, Thales, SAP, NCC, TfL, Cosmote, Net7).

The project work plan is completed with **Work Packages** on **“Training” (WP3)**, **“Dissemination and Outreach” (WP4)** and **“Coordination and Management” (WP5)**. WP3 is designed to oversee all the training activities in the network (schools, workshops, seminars). WP4 focuses on Dissemination and Outreach activities. It has therefore the twofold aim to create events to promote and exploit the results of the projects to the scientific communities and to reach out to the general public (high school students, undergraduates and teachers) in order to promote fundamental science and Big Data science and technologies. Finally, WP5 concerns the “**Coordination and Management**” structures of the network. Full description of the WPs is given in Table 3.1.

**Table 3.1: Description of Work Packages**

|  |  |  |
| --- | --- | --- |
| **WP Number**  | 1 | **Start Month – End Month** 9-45 |
| **WP Title**  | Research in fundamental physics |
| **Lead Beneficiary** | CEA |
| **Objectives**: Undertake research to answer fundamental questions about the forces and particles of nature, including the nature of Dark Matter, Dark Energy, Gravity and the Higgs sector. This research bridges several fields and experiments, with the aim of exploiting synergies between the different fields to enhance the research outcome via cross-fertilisation and the use of innovative big-data solutions as outlined in WP2, as well as by combining the research outcomes where appropriate. The aims are:(i) Analyse the LHC dataset delivered until end of 2018 to (a) search for Dark Matter signatures at hadron colliders, either discovering a dark matter candidate or setting world best limits on their production/decay; (b) measure Higgs boson couplings with high precision, to determine if the Higgs boson discovered at the LHC is the Standard Model Higgs boson and is responsible for the mass of all fundamental particles; (c) search for new physics in di-boson topologies and in Higgs production and decays, to either discover physics beyond the Standard Model, such as gravitons, or set world best limits on a range of models. // It would be good to add more explicit statements on how these objectives will answer the big fundamental questions in physics as has now been done for the ATLAS objectives(ii) Perform a large number of HPC simulations of binary neutron star mergers and develop efficient techniques to analyse a large volume of gravitational-wave data with a matched-filtering technique.(iii) Establish 3D data compression and Bayesian Hierarchical modelling methodologies for an efficient implementation of the analysis of 3D weak lensing data that Euclid will deliver.(iv) Define statistical procedures for the simultaneous extraction from data of large numbers of model parameters and nuisance parameters related to systematic uncertainties.(v) Combine Euclid dark matter sensitive data with other catalogues using novel methodologies for an efficient merging of information. Develop 3D dark matter mapping methodologies.// Add another objective on combining the research outcomes? Discussing results in workshop with theorists to identify how combined research outcomes can further our knowledge of the Universe? Here we should add the interactions among ESRs. The idea here is to start by asking each ESR to explain his/her research to all other ESR and then push them to find links and connections among research projects. The collections of seminar and the phd courses should help them on this point.(vi)  |
| **Description of Work and Role of Beneficiaries / Partner Organisations** // Some more mention of the different beneficiaries is needed and where their expertise lies is needed in this section.// This sentences “There will be little direct involvement of the partner organisations,… “ is not correct. INAF, INFN will be involved as mentors for EUCLID (INAF) and VIRGO and ATLAS (INFN). Also INFN will provide infrastructures. The tasks in this WP will mostly be carried out by the beneficiary organisations via their involvement in the ATLAS, LIGO-VIRGO and EUCLID experiments. There will be little direct involvement of the partner organisations, whose involvement in improving the research outcomes is via WP2, the outcome of which will directly feed this research. The basic underlying research is largely carried out on at experimental level, which is the combined research outcome then considered on a global level:ATLAS: Analysis projects in this WP aim at discovering new physics in a wide range of experimental signatures by looking at deviations from the Standard Model predictions in di-boson production or in Higgs boson couplings. Projects cover: single Higgs boson events (H->4l, H->bb), Higgs pairs (4b and 2tau2b) and di-bosons (WZ, WW, ZZ, ZH). Searches of new physics will be carried out looking for new resonances or for deviations from the Standard Model expectation in the Effective Field Theory framework. In all cases, a combination of data-driven techniques and simulation studies will be developed to optimize the selection criteria and to demonstrate a detailed understanding of the detector response and of the underlying physics. All studies will fully exploit the rich statistics of pp collisions that will be collected by ATLAS until the end of 2018. In case no new signals are found, exclusion limits will be set for a variety of physics models, including Dark Matter models. The projects are chosen to cover a wide range of new physics models, and at the same time to be complementary and provide crosschecks for possible new physics. The aim to cover such a wide range of studies requires expertise in many areas, this expertise is well spread among the different beneficiary and partners involved: UCL (trigger, b-tagging), UNIPI (Jet, Jet substructures), AUTH (… LIGO-VIRGO: The highly scalable 3D simulation code Einstein Toolkit will be used to perform high-resolution simulations of binary neutron star mergers with realistic equations of state in order to accurately extract the damping timescale of gravitational waves as well as of quasi-radial oscillations. Correlations between damping timescales and other physical parameters will be established leading to physically motivated analytic templates. Post-merger parameter estimation will be investigated using real data coming from the advanced LIGO-Virgo network. Observational constraints on the radius of neutron stars will be obtained and depending on the outcome the possibility for constraining alternative theories of gravity will be investigated. EUCLID: UCL, INAF and CEA teams a wide expertise exist on 3D weak lensing. The projects will develop 3D weak lensing in two ways: the development of 3D spin-wavelet reconstruction methods for 3D data compressions, and the development of Bayesian modelling. Both groups have world leading expertise in 3D lensing, at the strong and weak scale, and in the development and application of wavelet and Bayesian methods to cosmology. Based on the existing knowledge, pioneering studies will be performed on the application 3D spin-wavelets, of automated pipelines for cross-correlating the weak and strong lensing signal and to methodologies able to expend the sensitivity of the GLIMPSE algorithm will be extended to create 3D dark matter mapping over large areas of the sky.COMBINED: \\ Some discussion on the interplay/synergies between the various research outcomes and how we can jointly consider the outcomes to answer big-picture fundamental questions.**Task1.1** Optimization of the selection of physics objects (leptons, jets, taus, boosted objects, b-tagged jets…) and of tools for systematic uncertainty definition. Institutions contributing with complementary expertise: leptons (AUTH, UNILE, CEA), jets (UNIPI), b-jets and trigger (UCL) – involved ESR1-9.**Task1.2** Higgs pair production: definition of the analysis strategy for the Higgs pair production in the 4b (lead UCL ESR5) and 2tau/2b final state (lead UNIPI ESR1); the strong connection between the two ESR projects will be exploited for cross-checks and exchange of expertise on many common items (ESR1, ESR5).**Task1.3** Higgs coupling and exotic Higgs searches: definition of the analysis strategy for the associated VH(bb) production (lead UCL ESR6), H->4l (lead CEA) and ttH(bb) (lead CEA); strong interactions between the ESR projects will be exploited for crosschecks and exchange of expertise on the many common items (ESR6, ESR7, ESR8).**Task1.4** Di-boson searches: definition of the analysis strategy for the di-boson production searches in the leptonic (lead AUTH), leptonic and missing ET (lead AUTH), semi-leptonic (lead UNILE) and all hadronic (lead UNIPI) final states. Overlap between these projects and synergies with others will be exploited for crosschecks between the projects (ESR2, ESR3, ESR9, ESR4).**Task1.6** HPC simulations of binary neutron star mergers. Lead: AUTH. Institutions contributing with complementary expertise: GW analysis (UNIPI, INFN), equations of state (HITS), numerical techniques (Georgia Tech). ESR11 involved.**Task1.7** Implementation and use of a physically motivated analytic template for post-merger gravitational-wave emission. Lead: UNIPI, INFN. Institutions contributing with complementary expertise: template setup (AUTH, HITS), numerical techniques (Georgia Tech). ESR10 involved.**Task1.8** Development of 3D weak lensing methods using 3D spin-wavelet data compression methods (UCL, CEA), and Bayesian Hierarchical modelling (UCL, INAF, UNILE). UCL leads both these projects with expertise on wavelets (CEA) and Bayesian methods (INAF, UNILE) providing complementary expertise. The overall goal of these tasks is to measure dark energy properties with Euclid data.**Task1.9** Dark matter mapping and cross correlations. Development and application of cross-correlation methods between Euclid and multi-wavelength external data (CEA, INAF, UNILE), and development of 3D dark matter mapping methods for all-sky analysis (CEA, UCL). CEA leads both these tasks with expertise on external data (INAF, UNILE) and compressed sensing (UCL) providing complementary expertise. |
| **Description of Deliverables*** 1. Publication on development of cross-correlation and 3D mapping codes – M24

// Not clear what ‘public’ means at the end of this statement (publicly released)?**1.2** Testing of code and predictive forecasting made for 3D and cross-correlation public – M30**1.3** Application to Euclid first data of 3D mapping codes and cross-correlation – M40. **1.4** Publication on development of 3D-spin wavelet code and Bayesian codes to first EUCLID data – M40**1.5** Conference proceedings on the extraction of damping timescales through numerical simulations – M43**1.6** Publication on double Higgs boson production or derivation of cross-section limits in HH->4b, HH->2tau2b and their combination - M43**1.7** Publicationmeasuring the ttH coupling – M43**1.8** Publication onhigh precision measurement on Higgs coupling and new Higgs – M43**1.9** Publication on discovery of or world best limits on new physics searches in di-boson final states – M43**1.10** Publication or conference proceeding on the implementation of analytic gravitational-wave templates – M43**1.11** Report on how combined research output has furthered our understanding of the key outstanding questions in fundamental physics, the interplay between these results and what the next steps are to further enhance our understanding. // Remove from deliverables (do we really need a formal report on these items, which seems to add complexity for minimal gain)?// Yes, I think it will help to push the search for links and connection which is the most difficult part of the game… **2.3** Report on development of 3D-spin wavelet code and Bayesian codes – M24// Not clear how this is different to 1.2? Also not clear what ‘public’ means?**2.4** Testing code and predictive forecasting public – M30 |

|  |  |  |
| --- | --- | --- |
| **WP Number**  | 2 | **Start Month – End Month** 9-45 |
| **WP Title**  | Big Data in research and industry |
| **Lead Beneficiary** | UCL |
| **Objectives**: The aim of this work package is to facilitate knowledge transfer between different sectors by critically examining the challenges and use of big-data in both research and industry. This knowledge will be used to further the research output outlined in WP1, foster better links across sectors, ensure innovative advances in fundamental physics are exploited to enhance the economy and improve the skills/employability of the ESRs. Based on the research and secondment projects, the aims of this work package are to:(1) Acquire a critical view on the tools, frameworks, methods, challenges and innovations in big-data in both industry and research.(2) Identify novel/innovative solutions, areas where synergies exist or cross-fertilisation is possible both between and within different research fields and industry, that can address key challenges and/or enhance performance.(3) Implement enhancements identified in (1) and (2) in ESR projects carried out in WP1 and assess performance improvements.(4) Implement novel/innovative techniques identified in (1) and (2) to address big-data challenges or to enhance practices at the partner institutes.(5) Identify additional skills and techniques to be addressed in WP3 based on (1) and (2) where deficiencies have been identified. |
| **Description of Work and Role of Beneficiaries / Partner Organisations:**// More mention of the various partner organisations and where their expertise lies is needed in this section.The research and secondment projects will be used as benchmarks to acquire a critical view on the way the data is handled (storing, processing, monitoring, managing movement) and information is extracted (high-rate event selection algorithms for rare events, advanced data techniques statistical methods, multivariate analysis, machine learning methods, use of high performance computing) in different sectors. This will firstly focus on a survey of the available methods and tools, based both information acquired during the training, from the projects in the various research environments and from discussions with the partner organisations. This will focus on identifying both challenges and innovate/novel solutions. The aim is to identify areas where synergies and cross-fertilisation can be exploited, to ensure that knowledge transfer is maximised, to provide the most beneficial outcome in terms of enhancing the physics output, the impact on industry and the economy and the training of the ESRs. The beneficiaries through the supervisors will have a primary role to guide the ESR in their surveys and in identifying area of application in the research project. INFN, CERN, the IT-Department at UNILE, AUTH and SAS will provide the mixed IT expertise complementing the physics expertise of the supervisors to help in all the steps and in proposing tools/methods to be applied for the needed application. NET7, Cosmote, CERN, CMCC, … give input to this WP with the secondment projects.**Task2.1** Investigate the techniques used to manage and analyse data in ATLAS (ESR1-9), Virgo (ESR10, ESR11) and EUCLID (ESR12-15) (lead AUTH and in collaboration with UCL, UNIPI, CEA, UNILE, CERN).**Task2.2** Collect feedbackdetails of problems and solutions used in industry via partner institutes.**Task2.3** Identification of further research and secondment projects that can be enhanced from innovative solutions identified in T2.1 and T2.2.**Task2.4** Investigate the use of parallelized data processing, role of High Performance Computing (ESR10-15) (lead AUTH and in collaboration with UNILE, INAF, UCL, CERN).**Task2.5** Feasibility study to improve the analysis sensitivity with new analysis methods: Machine learning methods (ESR1-9), 3D-spin wavelet (ESR14), Bayesian hierarchical methods (ESR12).**Task2.6** Exporting of tools and methods from research environment to secondment projects and vice versa (lead AUTH in collaboration with partners), also ensuring the tools and methods are made publicly available, easily accessible and are well publicised.**Task2.5** Implementation of innovative solutions and assessment of performance. |
| **Description of Deliverables****3.1** Report on the handling and analysis of data in both the research (All ESR) and industry sectors, identifying challenges, innovative solutions and areas where synergies, knowledge transfer and enhancements are possible – M13**3.3** Report on implementation and impact of new techniques in the ESR projects – M30**3.4** Report on feed through of tools and methods both within and between, research and industry. Where appropriate new tools/techniques should be made publicly available and widely publicised – M35**3.5** Report on deficiencies in the training of PhD students with respect to big-data skills that should be addressed across fundamental physics – M35 |

|  |  |  |
| --- | --- | --- |
| **WP Number**  | 3 | **Start Month – End Month** 9-45 |
| **WP Title**  | Training |
| **Lead Beneficiary** | UNILE |
| **Objectives**: (1) To ensure timely delivery of workshops; (2) to setup a system to evaluate the effectiveness of the courses with feedback from ESRs; (3) to organize the full set of soft-skill courses, business data project seminars, technical and research presentations planned by the project; (4) to oversight secondments; (5) to ensure a timely development of the SW carpentry path; (6) to help identifying any need for additional training based on the output of WP2. |
| **Description of Work and Role of Beneficiaries / Partner Organisations**ESR will participate to many training events organized by the network as described in Section 1.2. All these events need to be organized timely to fulfil the tight schedule of the project. ESRs will also participate to one elective school on HPC or statistics. Aim of this WP is to identify the best schools available for our fellows. Secondments will also be overseen by this WP with the aim of ensuring they are planned at the best time for each researcher.**Task 3.1** Coordination of the training events. WP5 will be responsible for the timely organizations of the workshops. The AB will setup a programme committee for each event of the network. The committee will consist of WP5 leader, experts of the network and a local scientific responsible. This committee will be responsible of the scientific programme and of the timely delivery of the workshop. A local committee will take care of logistic aspects. **Task 3.2** Setup an event evaluation procedure and collect ESR feedback to be distributed to the SB and PO. **Task 3.3** Oversight secondments: an initial indicative secondment plan has been prepared. Once the ESRs are recruited and the ESR Supervising Teams (Section 3.2) are in place the secondment plan will be revised and optimised. WP5 leader will collect all proposals for modification and communicate the revised plan to AB and PC.**Task 3.4** Identify courses for elective schools and additional training in case of specific needs. WP5 leader will keep AB informed of the definition of elective courses and of additional requests. **Task 3.5** Ensure full publicity of the training material on the network website. |
| **Description of Deliverables****3.1** Report on the Big Data schools with feedback from the ESRs – M24**3.2** Report on effectiveness of secondment plan – M30**3.3** Report on SW carpentry full cycle effectiveness with feedback from ESRs – M40**3.4** Web-based repository fortraining material – M48 |

|  |  |  |
| --- | --- | --- |
| **WP Number**  | 4 | **Start Month – End Month** 6-48 |
| **WP Title**  | Dissemination and Outreach |
| **Lead Beneficiary** | AUTH |
| **Objectives**: (1) Communication to the general public of the project scientific and technological context; (2) educate fellows to communicate science and technique with effectiveness; (3) attract public, especially young students, to science; (4) disseminate the scientific results to different scientific communities. |
| **Description of Work and Role of Beneficiaries / Partner Organisations**This WP has a twofold aim of communicating to the public the scientific and technological challenges and to present the scientific results to a large scientific community. The WP6 leader is responsible for planning of the dissemination and outreach events that are described in Section 2.4. He/she is also responsible of the timely delivery of these events: any problems will be communicated to the AB and PC to identify corrective actions. The network will benefit of the large experience and technical tools of the GUARCO company, one of the network partner, which will also participate in defining the outreach programme. Particular attention will be put in highlighting the experience of the ESR work in both academic and non-academic environments.**Task 4.1** Provide website content and assist PO in drafting recruitment material.**Task 4.2** Draft a dissemination and outreach plan including events described in Section 2.4. The dissemination and outreach plan will be drafted trying to profit at best of the interaction with other EU project dissemination and outreach events of members of the network (SoBigData, GraWiton).**Task 4.3** Advertise the programme and the planned dissemination and outreach events. **Task 4.4** Dissemination of results in the scientific community at large and in particular in the two network conferences. |
| **Description of Deliverables****4.1** Dissemination and outreach activity plan public – M10**4.2/6.3/6.4** First Documentary public – M14/M24/M36**4.5/4.6/4.7** Annual reports on dissemination and outreach activities at annual meeting – M18/M30/M41**4.8** Scientific publication report from WP1-WP2 – M48**4.5** Dissemination and outreach material on Website - M48**4.6** Scientific conference report – M48 |

|  |  |  |
| --- | --- | --- |
| **WP Number**  | 5 | **Start Month – End Month** 1-48 |
| **WP Title**  | Management and Coordination |
| **Lead Beneficiary** | UNIPI |
| **Objectives**: (1) To ensure effective and timely implementation of the network project by coordinating and supervising the activities of all stakeholders and (2) to maintain communication with EU on the activity and financial status of the project. |
| **Description of Work and Role of Beneficiaries / Partner Organisations**A Programme Office (PO) will be setup at UNIPI and will serve as management office of the network as described in section 3.2. The PO will coordinate the day-to-day activity of the network including administrative, financial and communication matters. The main tasks of this WP are:**Task 5.1** Coordinate the installation of all the Boards and committees of the project in a timely manner;**Task 5.2** Make sure that all committees are well functioning and work according to the plan (see section 3.2 for role and aim of each function);**Task 5.3** Make sure that information distribution across and outside the network works properly at each stage of the project this includes the responsibility to have the project website up/running and up-to-date;**Task 5.4** Support WP7 and WP8 leaders in the organization of training, dissemination and outreach;**Task 5.5** Coordinate with doctorate offices at double-degree awarding University offices. |
| **Description of Deliverables****5.1** Project website up and running with general description of the project and ESR online application and registration system with a full description of posts opened – M1**5.2** Kick-of meeting at UNIPI – M1**5.3** Recruitment report with information on recruited ESR – M5**5.4** European Joint Degree Agreement (EJDA): Signed detailed agreement between each two universities awarding the double-degree specifying rules to obtain double PhD degree in the programme – M3**5.5** Declaration of conformity for each ESR – M6**5.6** Career Development Plan for ESR prepared – M9**5.7** Report on third year double-degree requirement fulfilment – M40**5.8** Final project report with all activities and results obtained – M48 |

**Table 3.1 b Deliverables List**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deliverable Number** | **Deliverable Title** | **WP No.** | **Lead Beneficiary Short Name** | **Type** | **Dissemination Level** | **Due Date** |
| 3.1 | Report on data managing structure and identification of critical points | 3 | AUTH | R | CO | M13 |
| 4.1 | Report on data analysis structure and identification of critical points | 4 | UNIPI | R | CO | M13 |
| 1.1 | Report on analysis strategies and ESR interconnections | 1 | CEA | R | CO | M18 |
| 3.2 | Report in ESR projects to proceed with data management optimization | 3 | AUTH | R | CO | M23 |
| 4.2 | Report in ESR projects to proceed with data analysis optimization | 3 | AUTH | R | CO | M23 |
| 1.2  | Report on novel approaches for events selection and statistical analysis | 1 | CEA | R | CO | M24 |
| 2.3 | Simulations with 3D-spin wavelet and Bayesian code (EUCLID) | 2 | UCL | R | CO | M24 |
| 2.6 | Publication on Simulation with cross-correlation and 3D mapping codes (EUCLID) | 2 | UCL | PDE | PU | M24 |
| 2.4 | Testing code and predicting forecast public | 2 | UCL | OTHER | PU | M30 |
| 2.7 | Testing code and predicting forecast made for 3D and cross-correlation public | 2 | UCL | OTHER | PU | M30 |
| 3.3 | Report on data management optimization | 3 | AUTH | R | CO | M30 |
| 4.3 | Report on data analysis optimization | 4 | UNIPI | R | CO | M30 |
| 3.4 | Report on cross-fertilization academy/non-academy for data management | 3 | AUTH | R | CO | M35 |
| 4.4 | Report on cross-fertilization academy/non-academy for data analysis | 4 | UNIPI | R | CO | M35 |
| 2.5 | Publication on 3D-spin wavelet and Bayesian code application to first EUCLID data | 2 | UCL | PDE | PU | M40 |
| 2.8 | Application of cross-correlation and 3D mapping codes to first EUCLID data | 2 | UCL | R | CO | M40 |
| 1.3 | Presentation of new results on double higgs production | 1 | CEA | PDE | PU | M43 |
| 1.4 | Presentation of new results on Higgs couplings and new Higgs | 1 | CEA | PDE | PU | M43 |
| 1.5 | Presentation of new results on ttH production | 1 | CEA | PDE | PU | M43 |
| 1.6 | Presentation of new results on new physics in di-bosons | 1 | CEA | PDE | PU | M43 |
| 2.1 | Report on extraction of damping timescale with numerical simulations | 2 | UCL |  R | PU | M43 |
| 2.2 | Presentation on implementation of analytic gravitational-wave templates | 2 | UCL | R | PU | M43 |
| ***Management, Training, Recruitment[[22]](#footnote-23) and Dissemination Deliverables*** |
| **Deliverable Number** | **Deliverable Title** | **WP No.** | **Lead Beneficiary Short Name** | **Type** | **Dissemination Level** | **Due Date** |
| 7.1 | Website up and running with ESR application and registration system | 7 | UNIPI | ADM | PU | M1 |
| 7.2 | Kick-off meeting in Pisa | 7 | UNIPI | OTHER | CO | M1 |
| 7.4 | EJDA signed | 7 | UNIPI | ADM | PU | M3 |
| 7.3 | Recruitment report with information on recruited ESR | 7 | UNIPI | ADM | PU | M5 |
| 7.5 | Declaration of conformity | 7 | UNIPI | ADM | CO | M6 |
| 7.6 | CDP ready | 7 | UNIPI | ADM | PU | M9 |
| 6.1 | Dissemination and outreach activity plan public | 6 | UNIPI | ADM | PU | M10 |
| 6.2 | First documentary pill on the project public | 6 | UNIPI | PDE | PU | M14 |
| 6.5 | Annual reports on dissemination and outreach activities at annual meetings | 6 | UNIPI | R | PU | M18 |
| 5.1 | Report on network schools with ESR feedback | 5 | UNILE | R | PU | M24 |
| 6.3 | Second documentary pill on the project public | 6 | UNIPI | PDE | PU | M24 |
| 5.2 | Report on effectiveness of secondment plan | 5 | UNILE | R | PU | M30 |
| 6.6 | Annual reports on dissemination and outreach activities at annual meetings | 6 | UNIPI | R | PU | M30 |
| 6.4 | Second documentary pill on the project public | 6 | UNIPI | PDE | PU | M36 |
| 5.3 | Report on SW Carpentry full cycle with feedback from ESRs | 5 | UNILE | R | PU | M40 |
| 7.7 | Report on third year fulfilment of double-degree requirements | 7 | UNIPI | R | CO | M40 |
| 6.7 | Annual reports on dissemination and outreach activities at annual meetings | 6 | UNIPI | R | PU | M41 |
| 5.4 | Training material on website | 5 | UNILE | PDE | PU | M48 |
| 6.8 | Scientific publication report for WP1-WP2 | 6 | UNLE | R | PU | M48 |
| 6.9 | Dissemination and outreach material completed on website | 6 | UNILE | PDE | PU | M48 |
| 6.10 | Scientific conference report | 6 | UNILE | R | PU | M48 |
| 7.8 | Final project report with all results and activities | 6 | UNIPI | R | PU | M48 |

**Table 3.1 c Milestones List**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number** | **Title** | **Related Work Package(s)** | **Lead Beneficiary** | **Due Date**  | **Means of Verification** |
| M3.1 | Definition of ESR projects suitable for Data Management framework optimization | WP3,WP1,WP2 | AUTH | M24 | Produce a document with list of ESR projects and proposed methods for framework optimization |
| M3.2 | New management framework implemented and performance assessment evaluated  | WP3,WP1,WP2 | AUTH | M40 | Optimized framework available to the network and document with performance evaluation produced |
| M4.1 | Definition of ESR projects suitable for Data Analysis framework optimization | WP4,WP1,WP2 | UNIPI | M24 | Produce a document with list of ESR projects and proposed methods for framework optimization  |
| M4.2 | New analysis framework implemented and performance assessment evaluated | WP4,WP1,WP2 | UNIPI | M40 | Optimized framework available to the network and document with performance evaluation produced |
| M1.1 | All particle physics ESR with analysis strategy designed  | WP1 | CEA | M18 | Report of all ESR to the second year network meeting  |
| M1.2 | Public results on Higgs with full LHC data-set  | WP1 | CEA | M40 | Presentation at conference and/or publication |
| M1.3 | Public results on new physics with full LHC data-set | WP1 | CEA | M40 | Presentation at conference and/or publication |
| M2.1 | Implementation of new EOS routines in 3D simulation code | WP2 | AUTH | M18 | Presentation at conference and/or publication |
| M2.2 | Application of analytic post-merger templates to GW detector data | WP2 | UNIPI | M40 | Presentation at conference and/or publication |
| M2.3 | Public results on dark energy physics with Euclid first data-release using 3D wavelets and 3D Bayesian modelling | WP2 | UCL | M40 | Report in terms of papers on application to data, presentation at conferences, and public release of code via repositories. |
| M2.4 | Public results on dark matter physics with Euclid first data-release using cross-correlation and dark matter mapping methods | WP2 | UCL | M40 | Report in terms of papers on application to data, presentation at conferences, and public release of code via repositories. |
| M5.1 | Big Data schools completed | WP5,WP3,WP4 | UNILE | M25 | Two schools completed with material uploaded to the website |
| M5.2 | SW carpentry path completed | WP5,WP6 | UNILE | M40 | Whole cycle of bootcamps up to delivering SW Bootcamp completed – info on website |
| M5.3 | Educational modules of the PhD programme completed | WP5,WP1,WP2 | UNILE | M40 | all courses from double-degree awarding institutions completed |
| M6.1 | Website in place | WP6 | UNIPI | M1 | Webpage online with initial material on EJD public |
| M6.2 | Dissemination and outreach plan completed  | WP6 | UNIPI | M4 | Detailed plan of dissemination and outreach activity ready |
| M6.3 | Conference on Big Data projects done | WP6,WP3,WP4 | UNIPI | M31 | Conference site and material available  |
| M6.4 | Conference scientific results done | WP6,WP1,WP2 | UNIPI | M42 | Conference site and material available |
| M7.1 | ESR recruited | WP7 | UNIPI | M6 | Name and cv of recruited ESRs on Website |
| M7.2 | All thesis proposal submitted to Academic Board | WP7 | UNIPI | M16 | Thesis subject identified and approved by AB with thesis summary on Website |
| M7.3 | Academic requirements for double-degree fulfilled  | WP7 | UNIPI | M43 | Green light received from double-degree awarding institutions |

Table 3.1d Individual Research Projects

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR1*** | **Host:*****UNIPI*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:*****1.1, 1.2, 1.3, 3.1, 3.4, 4.1, 4.4***  |
| **Project Title and WP(s) to which it is related**: New physics in HH->2b2τ with ATLAS (WP1, WP4)  |
| **Objectives:** The fellow is expected to play a leading role in the search for new physics through Higgs-pair production in the 2b2τ final state with the ATLAS data. The non-resonant HH production (Higgs self-coupling) is a crucial process for the Standard Model Higgs, however the predicted cross section is too low to be measured at the LHC until now[[23]](#footnote-24). With the new LHC run, at higher energy and luminosity, the HH sensitivity will be greatly increased. The HH->2b2τ topology, together with the HH->4b decays (ESR5) is one of the most promising channels in this search. The tau-lepton pairs will be searched for both in the leptonic-hadronic (lep-had) and full hadronic (had-had) final states. In both cases, it will be vital to achieve efficient tau identification together with a strong rejection against light quark jets to suppress the backgrounds and the fellow will work on implementing and optimizing advance multi-variant techniques for this purpose. Optimal b-jet identification is also instrumental for the sensitivity of this channel and synergies and complementarities with ESR5 and ESR6 is fundamental. |
| **Expected Results:** (1) Implement of modern machine learning techniques for optimal signal-to-background discrimination. (2) Publish results of ATLAS HH->2b2τ, in combination with HH->4b, including interpretation in various theoretical models. (3) Publish phenomenology studies of this channel for future prospects at even higher luminosities. |
| **Planned secondment(s):** Transport for London (TfL): 4 months – Optimization of London Underground Asset Maintenance Strategy. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR2*** | **Host:*****UNIPI*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:*****1.1, 1.2, 1.6, 3.1, 3.4, 4.1, 4.4*** |
| **Project Title and WP(s) to which it is related**: Search for new physics in events with missing transverse energy and a hadronically decayed heavy boson (WP1, WP3, WP4).  |
| **Objectives:** The fellow will enter the very challenging ATLAS community to play a leading role in the search of new physics in events with missing transverse energy and a boosted boson decayed hadronically that is reconstructed as a single “FatJet”. This signature allows to search both for candidate Dark Matter particles produced in association with a Higgs boson (H🡪bb) and for new resonances decaying in WZ/ZZ (🡪qqvv) bosons predicted by many new physics models (Heavy Vector Triplet, Gravitons, W’, etc.). The fellow will be involved in the development and optimization of the FatJet reconstruction technique, a new and very interesting jet reconstruction method that has greatly advanced thanks to the very high boosted regime reached at the LHC. In particular, high performance W/Z and b-jet tagging of the FatJet will be required to effectively reduce the background in the two searches. The fellow will be required to optimize the trigger algorithm to be used for this signature. The fellow will be required to optimize the analysis framework (data access, monitoring of grid jobs, network access…) in order to obtain a highly efficient and fast turn-around of the analysis, a must in an environment where fast feedback is often required to meet the challenging analysis schedule. The fellow will focus on the use of advanced analysis techniques to optimize FatJet tagging techniques. Lastly, work will be dedicated to the search for new physics or setting of limits using the full statistics data-set collected up to 2018. This will lead either to new physic discovery or to setting the world best limit on many new physics models. Synergies with ESRs 1, 4, 5, 6 and 9 will be of great benefit. |
| **Expected Results:** (1) Design of a coherent analysis framework to obtain fast analysis turn-around (2) Optimization of FatJet techniques using cutting edge analysis and statistical tools with publication of results (3) Publication of final results of the analysis on new signatures searches in events with MET plus hadronically decayed massive bosons.  |
| **Planned secondment(s):** Thales: 4 months - Tools for Big Data services. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR3*** | **Host:*****AUTH*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.5, 3.1, 3.2, 3.4***  |
| **Project Title and WP(s) to which it is related**: Search for new physics in the final state with two leptons and missing energy with ATLAS (WP1, WP3, WP4)  |
| **Objectives:** Search for the existence of an additional heavy Higgs boson particle decaying to a pair of Z bosons, with the final signature of two leptons and missing energy. This is a challenging search since it requires a deep understanding of the missing energy in the event and at the same time provides the means to search for possible existence of dark matter signatures at the LHC. The fellow will play a leading role to introduce new analysis techniques to optimise the signal extraction and background suppression. To this end, the network’s training programme will help the fellow to introduce modern multi-variant techniques. The work will benefit from the synergies and complementarities with the work of ESR2. |
| **Expected Results:** (1) Develop and implement multivariate techniques for data analysis. (2) Publish of ATLAS results on the 2leptons + missing energy channel (3) Publish limits or indication for dark matter signatures.  |
| **Planned secondment(s):** Thales: 4 months – Securing IoT: data protection in new landscape |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR4*** | **Host:*****AUTH*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.6, 3.1, 3.4, 4.1, 4.4*** |
| **Project Title and WP(s) to which it is related**: Search for Anomalous Gauge couplings and Exotics (WP1, WP3, WP4) |
| **Objectives:** The objective of the project is to search for New Physics indirectly through anomalies in the couplings of the gauge bosons and directly by looking from deviations in multi boson mass distributions. To this end it is important to measure the production of multi-boson final states and search for deviations from the Standard Model (SM). To precisely compare the experimental measurements for multi-boson production to SM expectations, clean and low background signatures are necessary, thus the leptonic decays of the di- or tri-bosons are selected. New particles decaying to vector bosons will also show up as excess or deviations in mass distributions of the final state leptons or as anomalies in vector boson scattering. It is important to measure both triple and quartic vector boson couplings as each of them point to different operators in the Effective Field Theory Lagrangian. Recent theoretical studies have shown the synergies between the anomalous Triple gauge boson couplings and the Higgs couplings to boson and fermions. Sophisticated statistical methods with multi-dimensional likelihood functions with their correlation matrices and data combination techniques will be developed. This research project will greatly benefit from synergies and complementarity of the work of ESR7 and ESR9. |
| **Expected Results:** (1) Optimize with modern multivariate techniques the signal selection. (2) Extract the background with data-driven methods. (3) Develop the necessary statistical tools for multi-dimensional likelihood functions. (4) Incorporate in the fits the Higgs couplings in order to obtain the final results for anomalous Triple gauge couplings. |
| **Planned secondment(s):** CERN: 4 months – Real time technologies for large scale analysis using Cloud deployments |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR5*** | **Host:*****UCL*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.3, 3.2, 3.3, 3.4*** |
| **Project Title and WP(s) to which it is related**: New physics in HH->4b with ATLAS (WP1,WP3)  |
| **Objectives:** The fellow will play a leading role in the search for new physics through (resonant and non-resonant) Higgs-pair production in the 4b final state, with the data from ATLAS at the LHC. The HH->4b topology has been shown to have great sensitivity in LHC’s low energy data, and with the increased energy and luminosity in the coming years, as well as with novel ideas implemented in the event selection to optimize the signal-to-background discrimination, 10-100 times higher sensitivity can be achieved. Given that 4b is a fully hadronic final state, suppressing the large multi-jet backgrounds will be vital for achieving the ultimate signal sensitivity. The network’s training programme will help the fellow to develop modern multi-variant techniques for background suppression. In addition to the inclusive search, the fellow will study also some more exclusive topologies, which have been little (or never) explored to date, such as the presence of two forward jets (the so-called VBF topology), or the presence of missing transverse energy as predicted in the context of the theory of Supersymmetry. The latter may indicate the presence of neutralinos, a prime particle Dark Matter candidate. The work will benefit from the synergies and complementarities with the work of ESR1 and ESR6. |
| **Expected Results:** (1) Implement modern machine learning techniques for optimal signal-to-background discrimination. (2) Publish of ATLAS HH->4b results and interpretation in various theoretical models. (3) Publish phenomenology studies for future prospects of this channel. |
| **Planned secondment(s):** Semblent: 4 months – Efficient export generation of relational data |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR6*** | **Host:*****UCL*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.4, 4.1, 4.2, 4.3, 4.4*** |
| **Project Title and WP(s) to which it is related**: New physics in VH(bb) with ATLAS (WP1, WP4)  |
| **Objectives:** Despite the discovery by ATLAS and CMS of a particle at 125 GeV that is so far consistent with a SM Higgs Boson, the dominant decay mode of the SM Higgs to b-quarks, which is predicted to happen 58% of the time, has not yet been observed. Many BSM models predict that the coupling of b-quarks to the Higgs will be altered by new physics, so as well as being an essential channel to study the properties of the Higgs, it can also be an important probe for new physics. The fellow will focus on studying the VH(bb) channel, the most sensitive channel in which to discovery H(bb). One of the most important aspects of this analysis is b-jet identification (b-tagging). The fellow will investigate using the latest machine learning techniques to boost the performance of the multivariate b-tagging algorithms deployed at ATLAS. The enhanced versions of this tool, coupled with state-of-the-art statistical analysis and machine learning techniques, will then be used to analyse the full Run 2 dataset and to make the world’s most accurate measurement of the coupling of Higgs to b-quarks using the VH(bb) channel. Combined with other channels this will either confirm the SM like behaviour of the Higgs or provide proof of beyond the SM behaviour. The VH(bb) analysis will then be converted to make a precision cross-section measurement of the dominant V+b-jet background making use of the techniques and tools developed previously. This work has synergies with and complements ESRs 1, 5, 7 and 9. |
| **Expected Results:** Use state-of-the-art machine learning and statistical analysis techniques to: (1) maximise the b-tagging performance and to publish a paper on these improvements; (2) publish the world’s leading measurement of VH(bb) and an updated Higgs coupling combination result; and (3) make a precision measurement of the V+b-jet background.  |
| **Planned secondment(s):** *NCC: 4 months –* Network activity data management and online analysis |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR7*** | **Host:*****CEA*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.4, 3.1, 3.4, 4.1, 4.4*** |
| **Project Title and WP(s) to which it is related**: New physics in H->ZZ->4l channel with ATLAS (WP1, WP4)  |
| **Objectives:** One of the main challenges in particle physics is the extensive study of the properties of the newly discovered Higgs boson at the LHC. A pre-requisite to search for new physics Beyond the Standard Model (BSM) is to precisely measure the couplings of the Higgs boson. The process that will be used for this study is the decay channel of the Higgs to 4leptons, an extremely low background process, well suited for measuring these couplings. The fellow will carry out a comprehensive research programme to look for a hint of BSM in the Higgs sector via Higgs-boson precision measurements in the framework of Effective Field Theory (EFT). The EFT has many parameters that should be all considered simultaneously, presenting a real challenge from the point of view of signal simulations and data analysis. The network’s training programme will help the fellow to develop and implement new statistical tools and new techniques of data analysis. The results combined with other precision measurements will be transformed to a powerful tool to confirm or refute the presence of new physics at the TeV energy scale. This work will benefit from the synergies and complementarities with the work of ESR3 and ESR4. |
| **Expected Results:** (1) Develop and implement new statistical tools for data analysis. (2) Publish of ATLAS H->4leptons coupling measurements within the EFT framework. (3) Publish BSM limits or indication for new physics at the TeV scale. 4) Constraint the parameter space for anomalous Triple Gauge Couplings.  |
| **Planned secondment(s):** CERN: 4 months – Tools to support unsupervised machine learning on scientific or industrial data |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR8*** | **Host:*****CEA*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:*****1.1, 1.2, 1.5, 3.1, 3.3, 4.1, 4.3*** |
| **Project Title and WP(s) to which it is related**: Search for boosted ttH (H🡪bb) with ATLAS (WP1, WP4)  |
| **Objectives:** The fellow will participate in the search for the production of the Higgs boson in association with top quarks using the ATLAS data at the LHC, with a subsequent decay of the Higgs boson to a pair of b-jets. This process is essential in testing the Standard Model in the Higgs sector, since it is the only one giving direct access to the top quark Yukawa coupling. The Standard Model predicts that this process should be observable during the LHC Run 2. The complexity of the event topology requires sophisticated multivariate analysis techniques in two areas: (1) the identification of b-jets and (2) disentangling the ttH signal from the tt+jets background. The fellow will participate in both aspects, focusing on events with high transverse momentum (*boosted* topology). While this topology is promising thanks to the LHC increased centre-of-mass energy of 13 TeV, it requires dedicated optimization to cope with the high particle density specific to this topology. This work, particularly on optimizing the b-jet identification, has many synergies with the work of ESR5 and ESR6, and the collaboration between the fellows in the network will be mutually beneficial.  |
| **Expected Results:** (1) Observation of the ttH process in the boosted regime. (2) Ultimate measurement of the top Yukawa coupling as a test of the Standard Model. |
| **Planned secondment(s):** Transport for London (TfL): 4 months – Predicting and preventing train failures on the underground |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR9*** | **Host:*****UNILE*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***1.1, 1.2, 1.6, 3.1, 3.4, 4.1,4.4*** |
| **Project Title and WP(s) to which it is related**: New phenomena in multi boson final states at ATLAS (WP1, WP3, WP4) |
| **Objectives:** Several theory-motivated models of physics beyond the SM assume new symmetries, which might lead to resonant production of pairs of gauge bosons. This research programme will investigate the semi-leptonic final states using the modern techniques of vector boson tagging via the reconstruction of a FatJet (J) from the hadronic decay, which enhances the reconstruction efficiency in the regime of high momentum, while the leptonic decay of the other boson will be used for an easy trigger strategy. The final states llJ and lJ+missing energy will be considered. A discrimination of Z boson decays to b-quark pairs will be attempted by applying b-tagging techniques to the FatJet, with the aim of disentangling WW, WZ and ZZ production, thus enhancing the sensitivity to new processes affecting selectively one of these channels. Robust techniques are needed for background suppression, for vector boson identification based on the FatJet features and for heavy-flavour. Therefore, novel methodologies can be fruitfully applied to this analysis. The work will benefit from the synergies and complementarities with the work of ESR2, ESR4, ESR5 and ESR6. |
| **Expected Results:** (1) Implementation of modern machine learning techniques for optimal signal-to-background discrimination for FatJet with b-quarks and (2) documentation of the identification performance obtained (3) Publication of ATLAS results on VV production with interpretation in various theoretical models |
| **Planned secondment(s):** ENEL: 4 months – Energy Management for trading department |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR10*** | **Host:*****UNIPI*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.3, 2.4, 3.1, 3.2, 3.3, 3.4*** |
| **Project Title and Work Package(s) to which it is related:** Post-merger GW emission in binary neutron star (BNS) coalescences (WP2, WP3) |
| **Objectives:** ESR10 will gain familiarity with the existing data analysis pipeline for BNS coalescence detection used in the advanced LIGO-Virgo collaboration. Starting from this baseline, the main objective is the study of the gravitational wave signal coming from the merger and post-merger phase. This will make use of physically motivated analytic templates with input from numerical simulations. Post-merger parameter estimation will be investigated using real data coming from the advanced LIGO-Virgo network. Efficient techniques will be developed for analysing a large volume of data with a matched-filtering technique and for placing observational constraints on the radius of neutron stars and alternative theories of gravity. |
| **Expected Results:** (1) Efficient algorithms for the determination of EOS parameters, validated with software injections in real data. (2) Estimate about possible physical constraint on Equation of State (EOS) and alternative theories of gravity. (3) Actual constraints on EOS and possibly alternative theories of gravity, upon successful detection. |
| **Planned secondment(s):** * mentoring at INFN Pisa (G. Cella), HITS (A. Bauswein), GTECH (J. Clark)
* ENEL 4 Months - Smart City Malaga - Consumption profile
 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR11*** | **Host:*****AUTH*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.1, 2.2, 3.1, 3.2, 3.3, 3.4*** |
| **Project Title and Work Package(s) to which it is related:** High-Performance-Computing simulations of binary neutron star mergers (WP2, WP3) |
| **Objectives:** ESR11 will make use of the highly scalable 3D simulation code Einstein Toolkit to perform high-resolution simulations of binary neutron star mergers with realistic equations of state on the PRACE network. A new, efficient method for handling tabulated, hot equations of state will be implemented. The damping timescale of gravitational waves and of quasi-radial oscillations will be extracted and correlations with other physical parameters will be established, in order to reduce the minimum number of parameters needed in physically motivated analytic templates of gravitational-wave emission.  |
| **Expected Results:** (1) Implementation of an efficient EOS module in the Einstein Toolkit. (2) A large archive of simulation results at difference resolutions covering the allowed parameter space of initial configurations. (3) Extraction of damping timescales of gravitational waves and reduction of parameters in analytic templates. |
| **Planned secondment(s):** * mentoring at INFN Pisa (G. Cella), HITS (A. Bauswein), GTECH (J. Clark)
* CERN 4 Months – Data Project within OpenLab to be better defined
 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR12*** | **Host:*****UCL*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.3, 4.2, 4.4, 5.1, 6.4, 6.5*** |
| **Project Title and Work Package(s) to which it is related:** Bayesian Hierarchical modelling for weak lensing |
| **Objectives:** We will use Bayesian Hierarchical modelling to create a probabilistic model of the full Euclid data set, including astrophysical systematic effects (in particular the unknown alignment of galaxies caused by close interactions), instrumental systematics effects (in particular a probabilistic model for charge transfer inefficiency in the Euclid CCDs), and measurement systematics. This builds upon recent preliminary studies in this field that treat the galaxy distribution as a series of 2D planes in distance (Schneider et al., 2014; Alsing et al., 2015), we will extend this to the 3D setting (Kitching et al., 2014). We will develop a model that respects the casual process through which the weakly lensed photons travel, which because it is based on a physical picture, is expected to be a more efficient approach. This builds upon the pre-cursor work in Viola, Kitching, Joachimi (2014) that defined the probability distribution of the measurement process, and Niemi, Kitching, Cropper (2015) that describes the extracting of prior hyper-parameters from weak lensing data. We will then apply this first to simulations to demonstrate the method, then to existing public data (including the CFHTLenS and KiDS data), and then to the Euclid data. |
| **Expected Results:** The methods created will first be applied to simulations with an expected article in which we present the formalism and demonstrate that the approach is feasible on data. There will then be an application to current data where we will present cosmological results on including a measurement of the dark energy equation of state as a function of redshift. The final application, and the most significant will be to Euclid data, as it accumulates towards the end of this PhD. This application will work with the Euclid consortium first to help identify scale-dependent systematic effects in the data, and then to measure the dark energy equation of state as a function of redshift. |
| **Planned secondment(s):** NCC2 4 months – to gain expertise in parallel programming architectures to speed up the calculations in preparation for the Euclid full data analysis |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR13*** | **Host:*****CEA*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.4, 4.2, 4.4, 5.1, 6.4, 6.5*** |
| **Project Title and Work Package(s) to which it is related:** 3D Dark Matter Mapping |
| **Objectives:** Underlying the link between weak lensing and the compressed sensing theory, we have proposed a completely new approach to reconstruct the dark matter distribution in three dimensions using photometric redshift information, and we have shown from simulations that we can estimate with a very good accuracy level the mass and the redshift of dark matter haloes, which is crucial for unveiling the nature of the Dark Universe (Leonard, Lanusse Starck, MNRAS, 2014; Leonard, Lanusse Starck, MNRAS, 2015). This new concept opens us a new way to analyse weak lensing data set, and could therefore be very important for the Euclid mission. However, two problems remain to be solved before it can be proposed for Euclid: the first one is the estimation of errors on the solution. To be useful, the method has to provide error bars, and a good approach would consist in integrating the method into a Bayesian framework. The second point is the computation time. Euclid survey will cover 15000 square degrees, which represent a huge amount of data. Recent proximal minimization techniques will allow us to develop a very efficient 3D tomographic reconstruction method in a High Performance Calculation (HPC) framework, such as the one proposed for the Euclid mission, the Euclid Science Data Center. |
| **Expected Results:** We expect the PhD student to develop our 3D dark matter matter mass map prototype into a method that can be applied in the framework of the Euclid space mission. The new method will provide error bars using a Bayesian framework, and will be optimized using the most recent techniques developed in the optimization field. The method will be validated on Euclid simulations and on real data, especially the CFHTlens survey. |
| **Planned secondment(s):** AUTH 4 Months – Intrusion detection and alerting service |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR14*** | **Host:*****UCL*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.3, 4.2, 4.4, 5.1, 6.4, 6.5*** |
| **Project Title and Work Package(s) to which it is related:** 3D data compression and wavelets for weak lensing |
| **Objectives:** The objective of this project is to construct 3D data compression processes for the Euclid mission, to optimise the measurement of the dark energy equation of state. The geometry of the Euclid data will be a 3D spherical setting where the underlying field is a spin-2 quantity (ellipticities of galaxies). To extract frequency-space (or scale-dependent) information from this data traditionally requires a Spherical-Bessel transform. However, this is sub-optimal when there are complex masks in the data, and frequency (or scale) dependent systematic effects. To mitigate this, we will develop a 3D spin-2 wavelet formalism building on the formalism described in Leistedt, McEwen, Kitching, Pires (2015). We will optimise the wavelet formalism for the case of dark energy measurements, and extend the approach to account for galaxy alignments and instrumental effects. We will then apply this first to simulations to demonstrate the method, then to existing public data (including the CFHTLenS and KiDS data), and then finally to the Euclid data. |
| **Expected Results:** The methods created will first be applied to simulations with an expected article in which we present the formalism and demonstrate that the approach is feasible on data. There will then be an application to current data, where scale-dependent systematic effects will be focussed on in particular CCD scale and field of view scale systematic effects. With these taken into account the analysis will present cosmological results on current data including a measurement of the dark energy equation of state as a function of redshift. The final application, and the most significant will be to Euclid data, as it accumulates towards the end of this PhD. This application will work with the Euclid consortium first to help identify scale-dependent systematic effects in the data, and then to measure the dark energy equation of state as a function of redshift.  |
| **Planned secondment(s):** Semblent 4 months – Auto query generation with Gremlin |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fellow:*****ESR15*** | **Host:*****UNILE*** | **Enrolment:*****Y*** | **Start date:*****Month 9*** | **Duration:*****36 months*** | **Deliverables:** ***2.4, 4.2, 4.4, 5.1, 6.4, 6.5*** |
| **Project Title and Work Package(s) to which it is related:** Combining the future multicolour sky: weak lensing from Euclid, gamma, X-ray |
| **Objectives:** The fellow will focus on the large-scale dark matter (DM) distribution that can be traced in a number of ways. Among the most powerful is weak lensing which has the advantage of not requiring any assumptions as the relation between luminosity and mass and/or hydrostatic equilibrium. The DM distribution generates cosmic shear, i.e. distortions in images of distant objects, but also gives rise to X-ray emission by the gas heated in the DM potential well and, possibly, to gamma-ray emission due to the annihilation of the DM particles. The study will include the modelling of the expected signals in X and gamma rays and the development of automated pipelines for cross-correlating the weak and strong lensing signal with source catalogues from X and gamma-ray space telescopes. |
| **Expected Results:** (1) Modelling of the shear maps signal (also using N-body simulations) and of the expected signal in X-rays and gamma rays. (2) Development of automated post-processing pipelines (using as input the output data of Euclid) for cross-correlating the weak and strong lensing signal with classes of sources in X-ray (by XMM-Newton, Chandra, Swift and, eventually, the next coming Xipe telescopes) and gamma-ray (by Integral and Fermi LAT observatories) surveys, possibly compatible with the Virtual Observatory.  |
| **Planned secondment(s):** AUTH 4 Months – Environmental monitoring and energy efficiency |



## 3.2 Appropriateness of the management structures and procedures

**Network organisation and management structure:** There will be a dual management structure, with one pillar dedicated to overseeing the implementation of the work packages, the **WP Management Board (MB)** and another one to oversee the academic aspects of the EJD, the **Academic Board (AB)**. The work of both boards will be supported by the **programme office (PO)** and overseen by the **Supervisory Board (SB)**. An overall scheme of the boards and committees (that are described in the following) is shown in Figure 3.2a. All details of the SB, AB, MB and PO including role, composition, and interactions are given in Table 3.2a.

**Joint governing structure:** The **AB** will appoint an **Admissions Committee (AC)** to oversee the joint admissions and one **Supervisory Team (ST)** for each ESR dealing with their selection, supervision, monitoring and assessment. The details of both bodies are outlined in Table 3.2b. The ST will be the first body to resolve any issues for the ESR. If the ST is unable to resolve the issue, it will be raised to the level of the AB, then to the SB, which will be the ultimate arbitrator. The ESRs can also raise issues via a representative on the SB, which provides an independent channel of communication. The ST will be responsible for ensuring that all institutional requirements are met and that the ESR’s CDPs are fulfilled.

Figure - 3.2a Network Management structure

Table 3.2a – Outline of Supervisory, Academic, WP Management Boards & the Programme Office

|  |
| --- |
| **Management body: Supervisory Board (SB)** |
| **Role** | Governing body: oversees all aspects of the project; ultimate arbitrator on all matters. |
| **Composition** | One member from each beneficiary and partner organisation; Chairs of the WP Management Board and Academic Board; one ESR representative, elected by the ESRs; the finance officer from the Programme Office will be ex-officio member. The composition of the SB will ensure that the views of all stakeholders are taken into account when decisions are made. |
| **Chair** | Elected by simple majority. |
| **Decisions** | Simple majority, the Chair has the deciding vote in a tie. |
| **Duration** | * M1-M48
 |
| **Meetings** | Kick-off meeting (M1), then face-to-face meetings held yearly (M8, M18, M30 and M41), during the Network’s Annual Meetings. Additional meetings may be held, if necessary. |
| **Interactions** | * MB reports to the SB on WP status, AB reports on academic aspects and the PO reports on financial and administration matters. ESR representative reports on any concerns or feedback from ESRs.
 |

|  |
| --- |
| **Management body: Academic Board (AB)** |
| **Role** | Implementation of the EJDA; Selection, supervision, mentoring and assessment of ESRs. |
| **Composition** | One member from each Ph.D. awarding and beneficiary institution. |
| **Chair** | Elected by the SB |
| **Decisions** | Simple majority, the Chair has the deciding vote in a tie. |
| **Duration** | * M1-M45
 |
| **Meetings** | Quarterly remote meetings and one yearly face-to-face meeting (M9, M18, M30 and M42). |
| **Interactions** | * ESR supervisory teams and Admissions Committee report to the AB. AB provides quarterly reports and makes annual presentation to the SB.
 |

|  |
| --- |
| **Management body: Work Package Management Board (MB)** |
| **Role** | Implementation of all WPs. Individual WPs will be overseen by **Work Package Managers (WPMs),** who are responsible for ensuring that the milestones and deliverables are met for each WP. |
| **Composition**  | Project Coordinator and WPMs with selected representatives from the partner companies. |
| **Chair** | Project Coordinator. |
| **Decisions** | Simple majority, the Chair has the deciding vote in a tie. |
| **Duration** | * M1-M45 (WP1-3); M1-M48 (WP4-5)
 |
| **Meetings** | Quarterly remote meetings and one yearly face-to-face meeting (M9, M18, M30 and M42). |
| **Interactions** | * MB sends quarterly reports to the SB and presents at annual SB meeting. WPMs interact with ESRs and relevant partners to ensure deliverables are met.
 |

|  |
| --- |
| **Management body: Programme Office (PO)** |
| **Role** | Responsible for ensuring that all project activities are being pursued in a synchronised and efficient manner and implements SB decisions; provides financial and administrative support; oversees the management of the intellectual property rights (IPR); handles all external communication. |
| **Composition** | Project Coordinator, AB Chair, SB Chair and a finance officer. Hosted at the Project Coordinator’s institute where suitable structures already exist. If needed, the PO may be staffed with an extra officer, who will be jointly and equally financed by all beneficiary institutes.  |
| **Duration** | * M1-M48
 |
| **Meetings** | As required, to prepare reporting documentation for the SB and the European Commission.  |
| **Interactions** | * Collates information from ESRs, AB, MB. Issues quarterly reports to SB on finances, ESRs and work packages. Communicates with all external bodies.
 |

Table 3.2b - Outlines of the Admissions Committee and the ESR Supervisory Teams

|  |
| --- |
| **Management body: Admissions Committee (AC)** |
| **Role** | Oversees the ESR recruitment. Sets the admissions criteria and ranks the ESR applicants. |
| **Composition** | AB, plus additional members from partner institutions to provide input from non-academic stakeholders. |
| **Chair** | AB Chair. |
| **Decisions** | Simple majority, the Chair has the deciding vote in a tie. |
| **Duration** | * M1-M8
 |
| **Meetings** | Meetings will be held as required. |
| **Interactions** | * Report on ranked list of candidates to SB and AB. Regular interaction with PO regarding ESR applications. Communication with degree awarding institutes via the PO.
 |

|  |
| --- |
| **Management body: ESR Supervisory Team (ST) one for each ESR** |
| **Role** | Monitoring, evaluating, supervising, mentoring and assessing the Early Stage Researchers.  |
| **Composition** | One for each ESR, comprising a supervisor from each double-degree awarding institution and mentors from secondment organisations. For formal assessments, the AB Chair will join the ST. |
| **Chair** | Elected by simple majority. |
| **Decisions** | Simple majority, the Chair has the deciding vote in a tie. |
| **Duration** | * M9-M45
 |
| **Meetings** | Quarterly meetings to evaluate progress. Annual assessment meeting with the AB Chair. More regular informal meetings with the ESR will also be held. |
| **Interactions** | * Quarterly reports to AB.
 |

**ESR Recruitment Strategy, Selection and Admission**: Given our past experience, we expect strong competition amongst high-calibre candidates for all the ESR posts. Emphasis will be put on recruiting the best possible candidates regardless of their background and following the guidelines in the European Charter for Researchers. The advertising campaign will include the following:

- Specially produced posters, announcements and e-mails will be circulated to the beneficiary institutes and other top research institutes worldwide.

- Use of websites such as the EU portal ([EURAXESS](http://ec.europa.eu/euraxess/index.cfm/jobs/index)), [jobs.ac.uk](http://www.jobs.ac.uk/), [findaphd.com](https://www.findaphd.com/) and [inspirehep](https://inspirehep.net/?ln=en), as well as the respective sites of the relevant experiments’ collaborations, and those associated to journals ([Physics Today](http://jobs.physicstoday.org/), [Nature](http://www.nature.com/naturejobs/science/), [Science](http://sciencecareers.sciencemag.org/) and [New Scientist](http://jobs.newscientist.com/en-gb/)).

- The positions will be advertised on social media accounts (for instance [Facebook](https://www.facebook.com/) and [Twitter](https://twitter.com/?lang=en)) of both the partner and beneficiary organisations.

All advertising will highlight the unique nature of the ESR projects, the high calibre of companies at which secondments will be carried out, the excellence of the research and the highly increased employability of graduates from this ITN project.

The posts will all be advertised shortly after the project has been accepted, before its official start date, to ensure there is ample time to advertise and fill all the posts. The aim is to have all posts filled by M3 in order to begin by M9. This should provide sufficient redundancy in case of difficulties filling any posts. To facilitate the early advertisement of the posts, the ITN’s website and an online application tool will be available a few months before M1.

The AC will define a joint admission and selection process and ensure it is unified across the network. A common set of supporting documents will be required, including curriculum vitae, statement of research interests, two letters of recommendation and a transcript of grades. All applications and supporting documentation will be required to be in English, as this will be the working language of the network. In certain circumstances, applicants may be required to provide an internationally recognised English language diploma certificate, in line with the regulations of the degree awarding institutes.

All applications will first be made to the network, where the PO will process them. The AB, with input from the AC, will then certify that the minimum requirements of the programme and those of both the degree-awarding institutes have been met. The applications will then be forwarded to all the relevant institutes.

The AB will produce a shortlist of candidates for each ESR post based on their credentials. The short-listed candidates will be interviewed by a selected subset of the AC with expertise relevant to the post. To ensure consistency, two members of the AB will attend all interviews. The full AC will meet to discuss the shortlisted applicants for each post and rank them based on the interview outcome and their credentials. The SB 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.

**ESR Supervision, monitoring and assessment:** The ST will monitor and assess regularly the progress of the ESR. The ST will produce quarterly reports for the AB and an annual report for the SB. The ESRs will write the first year progress report to be assessed by the ST along with the AB Chair. The ESRs must pass these assessments to continue in the project. A final assessment will be carried out by M43 to sign-off on the completion of the ESR project before they undertake their thesis examination. The AB Chair, along with the AB, will monitor the assessment of the reports to ensure consistency of standards across the network.

The supervisors will meet regularly with their ESRs (it is envisaged that this will be at least weekly) on a one-to-one basis to discuss progress on their project. More formal monitoring will also take place through quarterly reviews, based on the ESR’s Career Development Plan (CDP). These reviews will ensure that all institutional requirements are met timely, and progress is in line with the CDP. They will also give the opportunity to the ESR to highlight any issues related to their career development. A short appraisal, jointly drafted and signed by the ESR and ST, will be the outcome of these reviews. The ST will prepare an annual report for their ESR to be presented to the SB.

**Table 3.2c Implementation Risks**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Risk No.** | **Description of Risk** | **WP #** | **Risk Level** | **Proposed mitigation measures** |
| R1 | Delay in recruitment | WP1-2 | Low | Advertisement of posts begins before programme officially starts.Additional recruitment rounds will be held till all posts filled.* Flexibility for some ESRs to start in the project up to M13.
 |
| R2 | ESRs unsuitable to complete the research project | WP1-2 | Low | Strong supervisory structure in place.Assessment in M18 to identify such cases and replace candidates. |
| R3 | ESRs unable to complete Ph.D. within the duration of the network | WP1-2 | Medium | Strong supervisory structure in place to identify and remedy any such issues at an early stage.Commitment from all participants to support ESRs beyond the end of network until all of them have completed their degrees. |
| R4 | Shortage of secondments because of partner declination | WP2 | Low | We have already made sure that some of the large partners involved in the project could accommodate more secondments (Thales, CERN, etc.) with appropriate data projects. |

**Risk management:** Four main areas of risk have been identified and are described in Table 3.2c along with the risk level and the proposed mitigation measures.

**Scientific integrity:** Scientific integrity is of the outmost importance to this programme. While none of the proponents of this ITN have experienced any incidents of scientific misconduct in the past, we want to ensure that procedures are in place to deal with them. The SB will be the ultimate arbitrator of any cases related to scientific integrity. In addition, any case of financial misconduct will be brought to the attention of the PO, who will involve the relevant external agencies in the case of criminal misconduct.

**Intellectual Property Rights (IPR):** The network will adhere to all European and international agreements on the protection of intellectual property. The PO will ensure that all participants abide by the laws governing IPR, provide advice to participants on IPRs, be responsible for the management of any IPRs generated as part of the programme and implementing the rules set out in the CA. In the very unlikely event that a conflict emerges about IPRs, then, this will be discussed and resolved by the SB. Special seminars will also be organised as part of the annual meetings to promote IPRs to all the academic participants of the programme, in order to encourage the commercial exploitation of possible research spin-offs.

**Gender aspects:** Gender equality and the promotion of women in science are given a very high priority in this ITN. In particle physics and cosmology, women in general only hold ~20% of faculty positions. The gender composition of the management structure in this network significantly surpasses this number. The PC and four out of five beneficiary institute representatives are women. The consortium’s target is that at least one third of the ESRs are female and will exploit contacts to ensure a large number of applications from women. Our past experience has shown that the composition of the consortium itself is highly likely to attract more women candidates and we will make sure that all selection and interview panels have at least one female member. Profiles of the leading female researchers and fellows in the network will be highlighted on the ITN webpage to promote greater involvement of women in the field.

## 3.3 Appropriateness *of the infrastructure* of the participating organisations

The beneficiary organizations of BigDAPHNE can grant access to the national and international research facilities where the experiments are located. In addition, all the PhD awarding and beneficiary Institutions can host the researchers in high-standard premises, including libraries, on-site laboratories and offices for administrative support, providing state-of-the-art infrastructures needed for the ESR projects. Full details are reported in Table 5.1. In particular, all fellows will have access both to local clusters and to the largest computing systems available in Europe and the US (like the entire LHC Grid Computing infrastructure) for their training and research purposes. AUTH, UNIPI, and CEA-Saclay have projects running on PRACE Tier-0 systems, while all research teams (UNIPI, CEA, UNILE, AUTH, UCL) have also access to national or university-wide computer farms. Thus, the necessary computational resources of the project are fully secured.

A big asset of the BigDAPHNE participants is the proven ability of the research teams to collaborate, shown in past projects. To this end, the five public universities and research institutions (UNIPI, CEA-Paris-Saclay, UNILE, AUTH and UCL) have strong research teams in Particle Physics, Cosmology and Gravity, and already four of them (UNIPI, CEA, AUTH and UCL) had a very successful collaboration within the EU-project ARTEMIS[[24]](#footnote-25), a training network, which had a major impact in paving the way to the Higgs boson discovery at the LHC in 2012. Recently, the collaborations of UNILE with UNIPI and AUTH on Di-boson physics and on advanced techniques on fast, online data analysis and selection (FP7-IAPP EU-project FTK, grant agreement n.324318), are making pioneering advances in particle physics research and instrumentation.

For the GW data analysis, additional computing power is provided through the collaboration with the partner institutes INFN Pisa (infrastructure in Advanced VIRGO experiment, resources of CNAF computing center, Pisa computing center and EGO), GATECH (infrastructure in Advanced LIGO experiment) as well as HITS.

The non-academic and academic partners chosen for the training programme of the project have the necessary infrastructure, experience and strong desire to be actively involved throughout the duration of the BigDAPHNE project, and provide all the courses, summer schools and secondments addressing the needs for Big Data tools and analysis techniques. Details of the resources and facilities available at the premises of each partner institution are given in Table 5.2 along with a description of the involvement in research and training programs.

## 3.4 Competences, experiences and complementarity of the participating Institutions

**Consortium composition and exploitation of partners' complementarities:** The consortium consists of 5 beneficiary institutions AUTH, UCL, UNIPI, UNILE, CEA. The first four beneficiaries award PhD while CEA will enrol the recruited ESRs in the PhD school of the partner organization Université Paris-Saclay. Among them strong links and cooperation already exist; they provide the starting point for implementing with excellent efficiency the networking activities of the project. The ATLAS research projects are originating from synergies between several participants to BigDAPHNE already developed during the scientific collaboration within the ATLAS experiment. The Pisa group after a strong involvement in the physics of jets and gauge boson couplings has moved the focus of his interest to rare and exotic topologies with displaced secondary vertices from b or tau decays; they have strong commitments on the deployment of an ultra fast online selection of such kind of final states based on programmable electronics and associative memories. On this ground a strong cooperation already exists with AUTH in FTK project. The projects of ESR1, ESR2 and ESR5, focusing on Higgs physics in topologies with b-quarks and tau leptons and on searches for a signature of Dark Matter particles with an un-balanced b-jet system in the final state define a shared research between UCL and UNIPI. UNILE, AUTH and CEA had key and complementary roles in designing, constructing and operating the detectors of the ATLAS Muon Spectrometer, providing excellent momentum resolution and fast identification capability for muons which, along with electrons, are the most powerful and clean physics objects hinting to interesting processes. Muon reconstruction and trigger performance has been crucial for the discovery of the Higgs boson. The projects of ESR3, 4 and 7 are consolidating the collaboration of these teams on research topics where lepton reconstruction is the key ingredient to explore Higgs physics and look for new phenomena searching for tiny deviations of nature from the precise expectations theorists derive based on the Standard Model. The project of ESR9 will study in a different, semi-leptonic final state, the same process studied in the fully leptonic final state by ESR4 focusing on hints for exotic resonant production. The project of ESR6 will investigate Higgs physics starting from the experience gained in the study of the most abundant SM background for Z+H->bb, which was carried with 7 TeV data with first raw involvement of the UCL and UNILE teams. Many of the research projects of the ESR (1,2,5,6,9) rely on difficult tasks like b-jet identification and novel techniques for the reconstruction of highly boosted particles decaying hadronically. A common ground of expertise will be built on these topics in the network moving from the experience already existing at UCL.  Among the BigDAPHNE teams involved with the Euclid projects, UCL, CEA and INAF have already a strong cooperation on the items of the instrument and of the research methodologies, like the analysis of 3D weak lensing data with the 3D spin-wavelets method. Within the Euclid project UCL lead and CEA both lead aspects of the weak lensing data analysis pipeline and are already in very close collaboration on this aspect, meeting several times per year and in weekly or daily communication. Moreover, BigDAPHNE participants from the EUCLID community have already good collaboration links from previous experience on the CFHTLenS project, the largest public weak lensing data set, as well as the precursor experiments of Euclid DUNE and Space. The projects ESR12 and 13 consolidate these links beyond the data analysis pipeline of Euclid to the scientific exploitation of the data.

The experience of the UNILE team on strong gravitational lensing and the long-standing collaboration with Dr. Scaramella (INAF) are the ground for the development of the ESR14 and 15 projects shared between this team, joining now Euclid, and the CEA and UCL groups). The co-location of the cosmology and particle physics groups in UCL and CEA provide an unprecedented opportunity for the collaboration between CERN and Euclid on dark matter and dark energy physics. Finally, the AUTH group has strong links and complementary expertise with UNIPI/INFN on the Advanced VIRGO project, since UNIPI/INFN participates in the detector operation and the data analysis of the VIRGO experiment, while N. Stergioulas from AUTH is an elected member of the Executive Board of the VIRGO-EGO Scientific Forum, with an expertise in numerical simulations and characterization of gravitational-wave sources. Furthermore, collaboration with the Advanced LIGO project is established through the partnership with GATECH (J. Clark).

The INFN partner will provide complementary expertise for mentoring both on ATLAS and Virgo research projects.

The Software Carpentry Foundation, the AUTH IT-Center and CERN have long-standing experience with courses on software, computing and Big Data tools but with complementary characteristics. While Software Carpentry gives introductory courses aimed at a wide scientific community, CERN courses deal with specialized tools for particle physics research and AUTH IT-Center is expert in Grid and cloud computing. CERN, INAF, CNR and the non-academic stakeholders like Thales, ENEL, NCC, TfL, Semblent provide expertise and secondments in a very wide and diverse range of Big Data projects. CNR and UNIPI members that are also members of SoBigData project and CERN members also members of Openlab (a consortium of CERN and world-leading ICT companies, like ORACLE, INTEL, Rackspace) will also ensure expert training, secondments and good contacts with interesting stakeholders.

**Commitment of beneficiaries and partner organisations to the programme**: All participants are fully committed to the Joint Doctorate Programme. The network will ensure that the young researchers enrolled will follow a well-designed programme of training through research under highly qualified supervision and through a coherent plan of courses. All beneficiaries and partners have committed sufficient time for supervision and mentoring. The organizations where training events or workshops and conferences will take place are committed to ensure appropriate resources and arrangements for such events. Facilities at any given beneficiary institution will be open to all ESR of the project. The commitment of all beneficiary and partner institutions documented by letters in Section 7.

1. T=Provide Training; S=Host Secondment; M=Provide Mentorship; O=Expertise for outreach activities. [↑](#footnote-ref-2)
2. ATLAS <http://atlas.ch/>, Euclid <http://sci.esa.int/euclid/>, VIRGO <http://www.virgo-gw.eu/> [↑](#footnote-ref-3)
3. “The Data Revolution and Economic analysis” – May 2013 - <http://www.nber.org/papers/w19035> [↑](#footnote-ref-4)
4. “Data, data everywhere” – Feb.2010 - <http://www.economist.com/node/15557443> [↑](#footnote-ref-5)
5. “Big Data: The next frontier for innovation, competition and productivity”, McKinsey Global Institute, http://www.mckinsey.com/insights/business\_technology/big\_data\_the\_next\_frontier\_for\_innovation. [↑](#footnote-ref-6)
6. “Riding the wave: how Europe can gain from the rising tide of scientific data”, October 2010 report to EC of the High Level Expert Group on Scientific Data <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> [↑](#footnote-ref-7)
7. For example, in the words of Chris A. Mattmann, senior computer scientist at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena: “To get the best out of big data, funding agencies should develop shared tools for optimising discovery and train a new breed of researchers”, in “A vision for data science”, Nature v. 493 p. 473-475, 14 January 2013. [↑](#footnote-ref-8)
8. R. Brock and M.E. Peskin, Working group summary of “Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 3: Energy Frontier”, arXiv:1401.6081; R. Aleksa at al, “Input for the Strategy Group to draft the update of the European Strategy for Particle Physics”, http://europeanstrategygroup.web.cern.ch/europeanstrategygroup/Briefing\_book.pdf, Jan 2013. [↑](#footnote-ref-9)
9. L. Amendola et al., “Cosmology and fundamental physics with the Euclid satellite - Review of the Euclid Theory Working Group”, June 2013; arXiv:1206.1225. [↑](#footnote-ref-10)
10. A. Kashlinsky*,* [Astrophys. J. 823, L25 (2016)](http://dx.doi.org/10.3847/2041-8205/823/2/L25) [↑](#footnote-ref-11)
11. Schneider, M. D. et al, 2015, ApJ, 807, 87; Alsing, J, et al. 2016, MNRAS, 455, 4452. [↑](#footnote-ref-12)
12. Kitching, T. D, et al. 2014, MNRAS, 442, 1326. [↑](#footnote-ref-13)
13. Leonard, A., Lanusse, F., & Starck, J.-L. 2015, MNRAS, 449, 1146. [↑](#footnote-ref-14)
14. Leistedt, B., McEwen, J. D., Kitching, T. D., Peiris, H. V. 2015, arXiv:1509.06750. [↑](#footnote-ref-15)
15. Data Science at LHC 2015, <http://indico.cern.ch/event/395374/other-view?view=standard> [↑](#footnote-ref-16)
16. ITN-EID Founded by FP7, proposal n. 316596. [↑](#footnote-ref-17)
17. Founded by H2020-MSCA-ITN-2014, Grant agreement n. 642069. [↑](#footnote-ref-18)
18. Founded by H2020-MSCA-ITN-2015, proposal n. 675440. [↑](#footnote-ref-19)
19. [http://www.liis.it/mw/index.php/LiiS:E-Learning/Streaming](http://www.liis.it/mw/index.php/LiiS%3AE-Learning/Streaming) [↑](#footnote-ref-20)
20. 2016 Bertinoro School: https://agenda.infn.it/conferenceDisplay.py?ovw=True&confId=11680. [↑](#footnote-ref-21)
21. For ESRs enrolled at CEA. [↑](#footnote-ref-22)
22. E.g. advertising vacancies. The individual recruitments should only be listed in Table 1.2a [↑](#footnote-ref-23)
23. Phys. Rev. D92, 092004 (2015). [↑](#footnote-ref-24)
24. ARTEMIS, “Investigation of the electroweak symmetry breaking and the origin of mass using the first data of the ATLAS detector at the LHC” was under FP6, grant agreement n. 35657. [↑](#footnote-ref-25)