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]

TABLE OF CONTENTS

LIST OF PARTICIPANT

Consortium Member	Legal Entity Short Name	Academic (tick)	Non-academic (tick)	Doctoral Degrees (tick)	Country	Dept./ Division / Laboratory	Scientist-in-Charge	Role of Partner Organisation
Beneficiaries								
Universita` di Pisa	UNIPI	x		x	IT	Dep. of Physics and Computer Science	Chiara Roda	
Aristotelio Panepistimio Thessaloniki	AUTH	x		x	GR	Dep. of Physics and Computer Science	Chariclia Petridou	
Universita` del Salento	UNISA	x		x	IT	Dep. of Physics	Stefania Spagnolo	
University College London	UCL	x		x	UK	Dep. of Physics	Nikos Konstantinidis	
Commissariat a L'Energie Atomique	CEA	x			FR		Rosy Nikolaidou	
Partner Organisations								
THALES Service SAS	THALES		x		FR		Bernardo Resende	Provide training, host secondements
ENEL Ingegneria e Ricerca S.P.A	ENEL		x		IT		Matteo Masotti	Provide training, host secondements
Software Carpentry	Software Carpentry	x			US		Jonah Duckles	Provide training
European Organization for Nuclear Research	CERN	x			CH		Alberto Di Meglio	Organize and provide training, host secondements
Semblant	Semblant		x		UK		Harry Howard	Provide training, host secondements
Transport for London	TFL		x		UK		Ben Cooper	Provide training,host secondements
Universite Paris Saclay	UNIPS	x		x	FR		ROSY	Provide training
NCCGROUP	NCC		x		UK		Lloyd Brough	Host secondements
Istituto Nazionale di Astrofisica	INAF	x			IT		Roberto Scaramella	Organization of training, provide mentorships
Centro Nazionale Ricerca	CNR	x			IT		Fosca Giannotti	Organize and provide training
Istituto Nazionale di	INFN	x			IT		Giancarlo Cella	Organize and provide training,

Fisica Nucleare								provide mentorship
GUARCO Interactive	GUARCO		X		IT		Guido Francioni	Provide tools and expertise for outreach

.

Introduction

In the past two decades, the possibility to produce and collect data from a great variety of sources, from sensors to internet footprints left from personal activity, has produced a huge societal shift. The scale of this change 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[1]. The growth of interest around this huge volume of data – normally named "Big Data" – can be inferred by the rapid increase of the Google searches for the term "Big Data" in recent years, as shown in Figure 1. Already in the eighties, it was understood that the information hidden in this huge amount of data could be of extreme interest for the widest type of fields ranging from business to science, health care, security and policy.

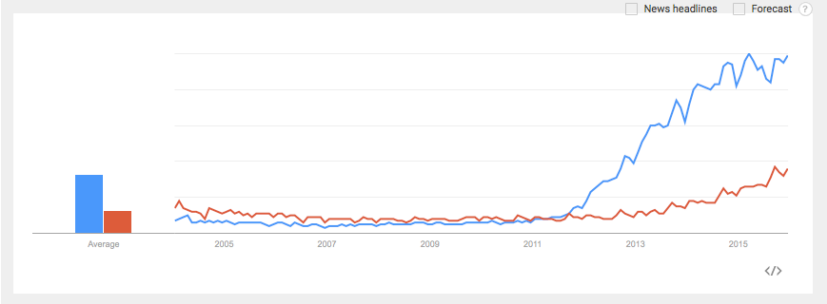

Figure 1 – "Big Data" (blue) and "Data Science" (red) Google search trends as a function of time over the past decade from https://www.google.com/trends/.

The ability to collect, store and analyze this ocean of data in order to extract useful insights has therefore become a very precious skill. Finding people able to develop the tools suitable for extracting this information and with the talent to put them in application in companies or academic institutions is presently a major challenge. Indeed, all observers agree that the abundance of big data coexists with a profound lack of data scientists, the professionals combining "the skills of computer scientist, statistician and narrator to extract the nuggets of gold hidden under mountains of data", proclaimed by The Economist as "the sexiest job of the 21st century"[2,3].

Some experiments in particle physics (ATLAS) and cosmology (EUCLID, VIRGO) have seen a very similar growth in data production and have a pivotal role in this trend. These experiments address fundamental questions on the structure of our universe, the origin of Dark Matter or the nature of gravity, by employing sophisticated analysis techniques to search for very rare signals hidden in huge amounts of data. The size of data that these types of searches require to collect (or simulate) and analyze is of the order of hundreds of PetaBytes per year, comparable to the yearly Facebook uploads. Students working in these research environments are therefore required to face challenges very similar to those found in big data handling or analysis tasks in industry.

Furthermore, exposing the young researchers to Big Data projects both from the academic and non-academic sectors, offers a very fertile ground to train them through an innovative path that enhances both the knowledge and experience in big data technology.

The BigDAPHNE project will complement a very solid training in fundamental physics with a structured training in data treatment and with secondments to institutions, which are either leaders in big-data technologies or power users of cutting edge data analysis strategies. Researchers will work in a multi-disciplinary international environment and will gain specific skills on Big Data technologies to become strong candidates for leading roles both in the academic and the industrial sector in Europe. BigDAPHNE will therefore contribute to populating the, so far, tiny community of EU data scientists, in response to the pressing need of science and society clearly identified by the European Community[4].

[1] "The Data Revolution and Economic analysis" – May 2013 - http://www.nber.org/papers/w19035

[2] "Data, data everywhere" – Feb.2010 - http://www.economist.com/node/15557443

[3] 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.

[4] "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

1 Excellence

1.1 Quality, innovative aspects and credibility of the research programme

Introduction

Particle physics research at accelerators and large-scale deep sky surveys, feeding observational Cosmology, are today strictly connected, providing insight one on the other. The unprecedented fast progress of the results collected over the past 50 years brought us to a turning point in our understanding of the Universe, where for the first time it is apparent that we do not know what 95% of the Universe is made of. This unknown consists of two components named, to partly reflect our ignorance of them, dark matter and dark energy, 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.

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 Higgs boson, a particle deeply connected with the theoretical foundation of the Standard Model (SM) of Particle Physics, has been discovered. Although theoretical calculations and experimental measurements have reached a very high precision, no prediction of the SM has been contradicted so far. Despite this great success, the SM describes only 5% of the content of the Universe and misses a dark matter candidate. Similarly, the Lambda Cold Dark Matter model (LCDM) provides a self-consistent description of cosmological parameters in very good agreement with all current observations. In spite of this, the density of dark energy inferred from current observations is 10^{60} times larger than that expected from particle physics. This is the largest discrepancy between theory and observation ever encountered in physics, a factor that cannot be reconciled without a fundamental re-evaluation of our understanding of the Universe. In addition, the texture of space-time, though remarkably well described by the 100 years old general relativity theory, presents a shortfall. Gravitational waves, a fundamental prediction of the theory, are still demanding to be experimentally established and to become a new powerful handle on cosmology. Large international research programs have been established to overcome the limits of current investigation in all these fields: the LHC is pushing forward the high energy frontier in particle physics with a program of data taking approved until 2023 and a high intensity extension (HL-LHC) planned for the following decade; a number of space and ground-based telescope programs 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 USA, and the corresponding European detector, Advanced VIRGO, will start operation before the end of 2016 near Pisa. **Based on these considerations it's clear that fundamental physics stands in a very interesting era.** The Joint Doctorate program BigDAPHNE intends to promote young researchers on projects within high perspective experiments in fundamental physics: the ATLAS[5] experiment at LHC, the EUCLID[6] space mission of the European Space Agency, and the Advanced VIRGO[7] experiment in Pisa. The interplay between the research domains of the project will be a strong asset of the network, which will bring in a scientific forum, for the first time, researchers acting in the foremost areas in each discipline. This will allow a common background on advancements to be injected in strongly connected areas of fundamental knowledge, that is typically not featured by standard research and training paths. The expected result of such work model is an enhancement of the vision in research trends and increase of the ability of sharing methodologies and tools across research domains. All of these projects share the challenges related to large data sets and tiny signals to be extracted from an overwhelming background. The BigDAPHNE project will foster the communication between the different research areas, by pursuing the development and exploitation of common or complementary advanced techniques to extract relevant features from large data sets and the deployment of common strategies to face at best the difficulties and exploiting the opportunities of Big Data. This

[5] http://atlas.ch/

[6] http://sci.esa.int/euclid/

[7] http://www.virgo-gw.eu/

plan is consistent with some key recommendations[8] to the European Commission and with a growing consensus in the academia about the need of a data strategy (for example, Chris A. Mattmann, senior computer scientist at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, says "To get the best out of big data, funding agencies should develop shared tools for optimising discovery and train a new breed of researchers"[9]).

Objectives

The overall goal of the project is to deliver a EJD program in fundamental physics and big-data science to train young researchers to the rigorous methodologies of scientific research enrolling them in world leading scientific experiments and empowering them with a focused training on advanced strategies for big-data handling. The main objectives are:

- Establishing a consortium where effective sharing of domain specific tools and strategies, particularly in the area of data mining, takes place leading to optimisation, enrichment of perspectives and improved discovery potential leading to progress in fundamental knowledge;
- Importing in the realm of rigorous scientific research the most promising trends on data analysis emerging outside academia or in research domains other than fundamental physics;
- Implementing a pioneering model of research and training network where the strong exchange of knowledge and practices between the academic research world and the non academic sector allows the new generation of scientists to figure out and undertake new paths to innovation with a significant impact for science and society.

Overview of the research programme

The European Joint Doctorate BigDAPHNE will pursue a research plan on fundamental physics contributing to three domains:

- **Scrutiny of the Higgs boson properties and searches for new phenomena** in proton-proton collisions, in the context of the ATLAS experiment at LHC;
- **Investigation of the nature of dark energy and dark matter** with the sky surveys from the EUCLID space mission of the European Space Agency;
- **Searches for signals of gravitational waves** with Advanced VIRGO in Pisa.

After the discovery of the Higgs Boson in 2012, the European and international scientific communities have deeply scrutinised the opportunities for new advancements in particle physics at present and planned accelerator facilities with the aim of optimising the balance of scientific impact and costs and risks of research programs. Strong motivations have been identified for exploring the TeV energy scale with a model independent approach and understanding the Higgs role. The outstanding position of LHC in the roadmap for High Energy Physics (HEP) clearly emerges from the relevant reports[10]. Euclid was selected as the second medium-class mission in the Cosmic Vision programme in October 2011. Redshift clustering and weak lensing are the methodologies that Euclid will use. 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, but by the rigorous treatment of systematic effects. Therefore Euclid stands in a leading position in the investigation of the history of the expansion of the Universe by complementing other powerful cosmological probes: cosmic microwave background and large scale structures, supernovae, abundance of galaxy clusters and strong lensing[11]. The Advanced VIRGO experiment, operated by the European Gravitational Observatory (EGO), is the most advanced (second-generation) laser-interferometer built in Europe with the aim of achieving the first direct detection of gravitational waves (known to exist due to observed evolution of binary pulsar

[8] "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

[9] Chris A. Mattmann, "A vision for data science", Nature v. 493 p. 473-475, 14 January 2013.

[10] 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, 13 Jan 2013.

[11] L. Amendola et al., "Cosmology and fundamental physics with the Euclid satellite - Review of the Euclid Theory Working Group", June 2013; arXiv:1206.1225

systems). The effort is in collaboration with the U.S. Advanced LIGO experiment: the experiments will operate jointly as a single large detector in order to maximise the possibility of initiating the era of gravitational-wave astronomy within the next 5 years. A design study for a 3rd-generation European detector (Einstein Telescope) has already funded by the European Commission, showing the great potential for fundamental physics in this field.

Among the technical innovations required by the research in these programs, the suite of Big Data related techniques and methods plays a very crucial role and deserves a common R&D effort.
Interaction with the research in the field of High Performance Computing and Information Technology (in terms of platforms and architectures) will be very important for the success of the program.

Table 1.1: Work Package (WP) List

WP n.	WP Title	Lead Beneficiary n.	Lead Beneficiary Name	Start Month	End Month	Activity type	ESR
1	Frontier Research in Particle Physics	5	CEA	9	45	Research	1-9
2	Frontier Research in Cosmology and Gravitation	4	UCL	9	45	Research	10-15
3	Big Data Handling Tools in Research and Industry	2	AUTH	9	45	Research/Training	1-15
4	Big Data Analysis Tools in Research and Industry	1	UNIPI	9	45	Research/Training	1-15
5	Training	3	UNISA	1	48	Management	1-15
6	Outreach and dissemination	4	UCL	1	48	Management	1-15
7	Management and coordination	1	UNIPI	1	48	Management	1-15

Research methodology and approach

The activities in the project are structured in work packages, listed in table 1.1. The first work package groups all the research items in the ATLAS experiments, with the goal of exploiting the large statistics of data at the frontier energy of 13-14 TeV that will be delivered by the LHC operation in Run2, started in 2015 and planned to last until the end of 2018. The second work package comprises projects leading to the studies of Dark Energy and Dark Matter at a cosmological scale within the context of the Euclid observatory and research projects related to the Advanced VIRGO program. The third and fourth work packages have a mixed flavour: they will implement the structured training of young researchers on data science and will enhance the research programs both in particle physics and observational cosmology and gravitation by reinforcing, updating and enriching the suite of data-handling and data mining tools for science. All the other work packages are aiming at optimising the implementation and impact of the project and will be described in details later in this document.

WP1 and WP2: The ATLAS, EUCLID and VIRGO research projects

The work model in High Energy Physics experiments at the frontier of energy and intensity, like ATLAS, critically depends on an efficient data strategy. LHC is delivering raw data at the rate of about 1 PB per second and the computing model of the experiments is based on the fundamental choice of reducing the data storage to about 1 GB per second by applying very selective hierarchical event reconstruction algorithms online within a total latency of 1s. The production of analysis-ready data is done through a complex suite of reconstruction algorithms, which need to access several sources of condition data and to propagate the core information embedded there in output metadata. Finally the production of physics results implies extensive simulation and recursive calibration and processing of data and simulation. The computing needs of this kind of model are addressed with distributed computing resources on the grid. High performance networking and computing, efficient data storage and data access architectures are mandatory for HEP research. Both reconstruction and physics result extraction have moved in the LHC era from simple cut-based techniques to machine learning (ML) techniques. Boosted decision tree, neural networks, multivariate analysis techniques are currently of wide use. However, the HEP toolkit is still limited to a subset of the full suite currently on the market; unsupervised learning techniques, deep machine learning, vision classification algorithms are still almost unexplored opportunities. These methods are typically cpu expensive and their efficient implementation requires innovative software approaches and languages (python, R-language, symbolic calculation languages) which are not yet familiar to the physics research environment. Other approaches, like wavelet analysis, have been rarely considered as a methodology for signal extraction in HEP. The HL-LHC program, where event complexity will increase by a factor of 20 and the data size collected for physics analysis will grow by a factor of 100 with respect to the present, is definitely an urgent case for widening the HEP methods and tools.

Addressing the most important questions in cosmology requires the best and most

comprehensive data analysis methods. Gravitational lensing is widely accepted to be such a method. When applied to the lensing of light from galaxies by the large-scale structures of the Universe, as a function of distance or look-back time, it enables to map the 3D dark matter structure of the Universe as well as its expansion history. This combination of lensing information and distance is known as 3D weak lensing and it is a field in which the Co-I's of this project are international experts. 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 the EUCLID data set. The Bayesian Hierarchical modelling is an innovative technique, recently pioneered in preliminary studies[12] and proposed as a tool for an unbiased interpretation of the EUCLID data in terms of dark energy[13]. In addition, recently 3D spin-wavelets have been developed that can be used to construct 3D weak lensing statistics from the data. Innovative 3D dark matter mapping method-ologies, extending the state-of-the-art compressed sensing GLIMPSE algorithm[14] (Leonard et al., 2014) 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[15] for cross-correlating the weak and strong lensing signal with source catalogues from X and gamma-ray space telescopes will be pursued to improve the combined physics reach in dark matter mapping.

Information about gravity in the strong interaction regime 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, which the Advanced VIRGO experiment might detect. An interesting possibility is that these signals can contain signatures of violations of General Relativity, in particular additional fields foreseen by extended gravity theories. These, in turn, can have an important cosmological role. The main tool for the detection and the study of merger events is the Wiener filter technique, which in case of unknown parameters amounts in the evaluation of a bank of templates large enough to cover the admissible signal manifold embedded in the space of possible time series. 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 techniques can 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.

BigDAPHNE aims at giving an important contribution to the program of exploring, exercising and consolidating new methods for data handling and data processing with the data that LHC and VIRGO will deliver in the forthcoming years, and in preparation of the challenges that EUCLID will face after 2020. The involvement of partner bodies of the network, both from the area of scientific research and from the non-academic sector, will inject complementary expertise.

WP3 and WP4: The Big Data inspired Work Packages

Research in particle physics and astronomy has been facing the problem of producing, storing, analysing large sets of data since a few decades. Several new paradigms have been established in the years through the computing models of big scientific collaborations. As a noticeable example, research in particle physics, at CERN with LHC in particular, has been the driving force for the development of distributed computing on the grid. 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 the recent years, the synergy with the non-academic ICT research took the form of the OPEN-Lab consortium, where R&D programs 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 about five years ago (see Fig. 1), built a case for the implementation of data cloud services from commercial providers: Amazon since 2002 and later Google since 2009. The LHC experiments are slowly migrating from pure grid to cloud-based computing models. 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 SoBigData project, lead by a consortium of

[12] Schneider et al., 2014; Alsing et al., 2015
[13] Kitching et al., 2014
[14] Leonard et al, 2014.
[15] Leisedt, McEwen, Kitching, Pires, 2015.

University of Pisa and CNR, is taking a prominent role in organising rigorous studies and training programs on data analytics. The directions of development of data processing models in the area of social studies, life science, marketing have so far little commonalities with the typical tools used in physics research, although it's likely that a contamination of the two fields might in fact generate benefits in both directions. Likewise the scientific communities working on sky surveys or in experimental particle physics share a surprisingly limited set of techniques in the practice of research. A recent workshop[16] held at CERN brought to a non-expert audience the interesting opportunities that might arise from a deeper communication between different forums. The BigDAPHNE project has the potential of contributing to closing the gap between technical competences on data treatment in different areas, including non-research applications. The main objective of these work packages is to inject in the training of young researchers specific deep competences on data handling technologies and data analysis. Trends in data management, distributed computing, relational and non-relational DBs, data mining techniques and algorithms, statistical inference will be the main focus of a dedicated training program issued by world experts from various fields. The application in the research program of each student of novel data analysis techniques borrowed by other applicative or research fields will be encouraged, aiming at standardising valuable tools in areas where their use is not yet common. A potential breakthrough would be the foundation of a open source library of algorithms implemented in a non-discipline specific manner for wide use in science and outside.

Strong input to the program of the WP3 and WP4 will come from the partnership of highly qualified institutions: OPEN-Lab, INFN, INAF, CNR and SSI, companies operating in diverse areas, like Thales, ENEL, NCC, TfL, where strong IT divisions manage large data analytics projects, and companies like Semblent which are providers of data analytics services to private clients. Existing relations with other non-academic stakeholders, like NetApp[17] and Pivotal[18], will be strengthened during the life of the project.

Collaborative nature of the research projects and interaction of the WPs

The three fundamental research areas hosted by the BigDAPHNE program are strongly connected. Search for Dark Matter direct production at LHC is the complement at microscopic level of the dark matter investigation at the cosmological scale; understanding Dark Energy requires the exploration of new force fields or modified gravity that are addressed with different methodologies in the search for new phenomena at the LHC energy frontier and in the search for gravitational waves of cosmological origin. Therefore the research program of BigDAPHNE features a strong inherent interdisciplinary and the collaboration of world experts will be exploited to enrich the interpretation of the outcome of the ESR research programs. In addition to that, the common challenge of extensive data collection and processing, shared nowadays by non-research domains, sets a global ground for profitable interactions in the network. The work packages related to big-data handling and mining of BigDAPHNE are a forum where knowledge exchange between several academic and non-academic partners is made concrete. By Sharing, consolidating and developing new methodologies and tools across the different disciplines and the research and practice in several non academic environments will enhance the toolkit for accelerating knowledge advancement and innovation and progress in society. The relations between the various work packages proposed are

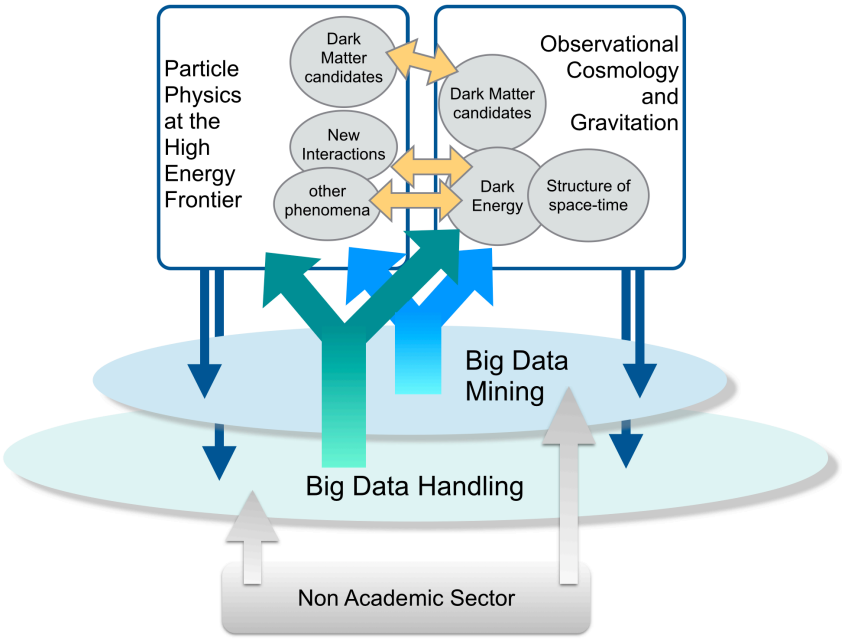

Figure 2: Interactions of the BigDAPHNE work packages.

[16] Data Science at LHC 2015, http://indico.cern.ch/event/395374/other-view?view=standard
[17] http://www.netapp.com/us/ and the SSICLOPS project https://ssiclops.eu.
[18] http://pivotal.io

schematically shown in Fig. 2. A rich program of workshops is proposed to help fostering the cooperation on the scientific and technical ground.

Originality and innovative aspects of the research programme (in light of the current state of the art and existing programmes / networks / doctoral research trainings)

The academic community in Europe is historically strongly involved in leading research activities both in HEP and in Observational Cosmology. Top level expertise in the traditional methods involved with these kind of research, from detector design and construction up to physics interpretation and model testing is wide spread. While detector design is widely felt as a research topic in physics and therefore included in research paths, the expertise in computing, software and data analytics design is typically treated as an accessory skill. 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. The novelty of the BigDAPHNE program stands in the choice of putting at the centre of the research program these crucial methodologies. The cross-disciplinary nature of this research area dictates the need of establishing synergies between various research disciplines and different sectors. BigDAPHNE faces this requirement by

1. Connecting two research domains in fundamental physics which are driving the effort in Big-Data technology progress, thus keeping a program well focused on common physics goals;
2. Establishing partnerships with leading institutes in IT and Data Analytics research;
3. Actively exploring the trends and opportunities emerging from the non-academic sector.

Recently innovative research paths have been traced at the border of physics research in the Icedeep EID and the HPC-Leap EJD programs (funded by EC under grant agreements …). While the focus of these projects is on high performance computing for scientific research, BigDAPHNE will pioneering the ground of data analytics and related software and computing architectures for the seek of science advancement.

1.2 Quality and innovative aspects of the training programme

Overview

The training programme provided by the network aims at educating the fellows:

→ to apply a solid scientific method to projects in different disciplines;

→ to lead projects in an international multi-disciplinary environment which also include the ability to present, document and defend the results of their work/researches with profit;

→ to master up to date competences in sustainability and scalability of software projects, "Big Data" management and analysis.

In order to reach these goals the training focuses on three pillars:

1. **PhD research projects** on pressing open questions about particles physics, cosmology and gravitation in international projects with world-experts from two institutions in two different countries; All the participant universities have well established doctorate programs in fundamental physics where the ESRs will be able to select and attend domain specific courses in particle physics, cosmology and gravitation to gain a solid preparation to their research activity.

2. 3-6 **months secondment at the premises of academic or non-academic partner institutions leaders in big-data technologies or power users of cutting edge data analysis strategies**; The researchers will work of projects of interest of the host institution, under a qualified supervision. They will either contribute to extract business information from data in the various fields (mobility, energy, internet security) or participate to R&D activities. The academic institutions offering internships (CERN and AUTH-Grid-center) are leaders in IT technologies and the Big Data managing methods and they have also close contact with the industry. This experience will add a valuable component to EJD Researcher's curricula that is missing in standard PhD paths.

3. structured courses, short workshops and seminars to build candidates with strong profiles on **good software practices, Big Data technologies, statistics, business mindset for data projects and communication**. These courses, specifically organised by the network, will be open to non EJD-researchers, and will complement the discipline specific background with the goal of providing skills that enable the fellows to be both more efficient in the research work and better formed for leading roles in the industrial or academic environment. In these training events the EJD fellows will benefit from the interaction with other young researchers in the same and in different disciplines, both at the hosting and seconded institutions. The network training activities are excellent opportunities for cross fertilisation between different fields where the sharing of expertise and tools will be vigorously fostered.

The structure of the training activities in the BigDAPHNE program will therefore 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 to a deep understanding of the thesis research field and to the ability to manage independently a research project. For all the universities of the network the doctorate PhD programs last three years. At the successful completion of the program the ESRs will gain a double degree.

Content structure of the doctoral programme

All the ESRs will be recruited at the same time. We assume to be able to have all ESRs recruited by mid 2017 in order to enrol them to the 2017 doctoral programs. The ESRs will therefore be able to participate to the network wide training courses at the same moment in their path. The events organised by the network are cross-disciplinary and are compulsory for all ESRs. The hosting institutes will take care of the local organisation. A program committee will be setup for each workshop that will always include the WP5 leader. This committee, in coordination with the Academic Board will be responsible for the workshop program. The workshop material will be published on the BigDAPHNE site. The training events will be also open to fellow external to the network. 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.

Network wide training events

The network wide training events are events 1 to 7 in table 1.2b The first course that EJD students will take (the Software Carpentry BootCamp – event 1) aims at providing the students with a solid understanding of how to compose and grow scientific software tools. The SW Carpentry foundation is leader in providing workshops on scientific software development and reproducible scientific methods to young researches of diverse disciplines. They have already experience in the training delivery for the ELIXIR UK community, the ELIXIR EU infrastructure project in biology[19]. Once skills have been developed, and in order to build capacity to spread the SW Carpentry experience we will have our students attending Instructor Training to become themselves SW Carpentry Instructors (event 6). This course will teach them modern pedagogical techniques for communicating scientific topics to scientists of diverse backgrounds. This educational path will be finalized with them acting as Instructors at one of the SW Carpentry BootCamps. This will happen between M24 and M43. In summary EJD students will attend a 2-day workshop, then work to develop their skills in collaboration with other EJD students over several month, then learn to the pedagogical techniques used to teach those workshops in a 2-day Instructor training course taught by Software Carpentry Instructor trainers, and then be mentored and supported by the Software Carpentry community to deliver a 2-day workshop to their own research community.

The Data Science education on Big Data will be provided with two two-week long schools. The first school (event 2) aism at providing and introduction to the Big Data world with particular enphasis on data and tools used in the most varied society environments. This school is organized by members of the

[19] ELIXIR project - https://www.elixir-europe.org/

SoBigData RI-H2020[20] project that are pioneer in Data Science and world expert in IT technologies for Big Data on in education in Data Science. The second Big Data School (Data Science in reasearch: focus on particle physics and cosmology - event 3) will be held at UCL and will focus on methods and technologies for managing and analysing data in the scientific fields of the network: accelerator particle physics, deep sky surveys and gravitational waves. The organization of the school is leaded by CERN members, which have a huge experience in IT school organization. The INAF, CEA and AUTH members will provide the additional expertise on technologies and methods used in cosmology. The EJD fellow's training will be completed with two-day workshop on Hadoop held in Thessaloniki and organized by AUTH-Grid center.

Two conferences held toward the end of the network will cover the technological aspects ("Big Data: a wide view from industry projects" – event 4) and the scientifica aspects (Conference on scientific harvest – event 6). Experts from companies and academic world will be invited to the two conferences and the ESR researchers will present their results in a wide international context.

Month 8 - Software Carpentry BootCamp – CEA

EJD fellows will learn how to use the UNIX Shell, Python and Git to better develop and manage scientific software and workflows through the development of software projects. The UNIX shell is taught with an emphasis on automation and repeatability of tasks at the command-line environment. Python is used as an example language to grow software based on small testable functions. Git is taught in order to teach version control and collaboration techniques for effective software development. At the end of the two days, trainee have a better idea about how these tools work together and some hands-on experience using them in supportive learning environment.

Month XX - Software Carpentry Instructor BootCamp - AUTH

SW Carpentry Instructor Training will be delivered by one of SW Carpentry's instructor trainers. Instructor trainers are senior instructors and mentor from the SW Carpentry community who teach a two-day workshop on pedagogy and instructional techniques. Using modern literature from research on teaching, ESRs will be guided to become reflective and introspective instructors, gaining skills and techniques for reducing cognitive load, treating teaching as performance art ~~and becoming reflective and introspective teachers~~.

Month 15 - Data Science: an introductory school – UNIPI

Courses at this school will be on: High Performances & Scalable analytics, SQL and noSQL databases, Data mining, Machine learning, Data Visualisation, ethic issues. Most of the courses foresee lectures and practical laboratories. At the end of the school students will be assigned a small project (on non-physics data) to be developed in the subsequent months. The projects will be finalised at a dedicated one-day workshop that will be held at the beginning of the second Big Data School (event 3).

Month 18 – Data Science in research: focus in Particle Physics and Cosmology – UCL

Courses will give insights on the framework used for managing and analysing the data in particle physics and cosmology. More emphasis will be given to data analysis spanning already used methods (examples for particle physics are: boosted decision tree, neural networks, multivariate analysis..; and for cosmology wavelets, Bayesian Hierarchical modelling..) and new tools on the market not yet exploited in particle physics (unsupervised learning techniques, deep machine learning, vision classification algorithms…) and in cosmology (3D-wavelets). The ESRs will profit from learning data handling and analysis methods used in research environments closely connected for what concerns the research objectives but very different for the technologies used. CERN members in collaboration with INAF, CEA and AUTH members will organise this school.

Month 30 – 2-day Workshop on Hadoop - AUTH

[20] SoBigData project – https://www.sobigdata.org/

Hadoop has become a must for Data Scientists in the company market a 2-day Workshop will give a introductory view of this important tool. Members of the AUTH grid and computing centre are experts in Hadoop and have already delivered courses on this framework. AUTH will provide the necessary infrastructure for hands-on laboratories.

Month 30 – Big Data: a wide view from the industry world – AUTH

Taking advantage of all the connections built during the network we will organise a conference where Big Data projects in society will be presented. ESRs will present their secondment projects in this venue.

Month 41 – Conference on scientific results – UNISA

In this conference emphasis will be given to the output of the ESRs research projects. New ideas, connection between particle physics and cosmology and tools used in research projects will also be discussed in a wide scientific community.

During the annual network meeting or during the network schools, training devoted to transferable skills will be organised. This will include training in presentation skills, scientific writing, CV writing. Many institutions in the network have expertise on organisation and delivery of these type courses (INFN, CEA, UCL…) 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). 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 enterpriseFor this we will both involve members of the network partners and other industrial stakeholders.

Role of non-academic sector in the training programme

The deep participation of members from the non-academic world to the organisation, supervision and training of the fellows is one of the distinctive features of this network. The industrial partners of the network are either word-leading developer of Big Data tools or users of cutting edge technologies, extracting profit valuable information from their data in the most varied fields: mobility, energy, internet security... they 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 interaction with the non-academic sector aims both at fostering the reaserchers's entrepreneurship and at improving the exchanges between the non-academic and academic environments.

Table 1.2 a - Recruitment Deliverables per Beneficiary

Researcher No.	Recruiting Institution Institution A	Institution B	Planned Start Month 0-45	Duration (months) 3-36
1.	UNIPI	UCL	9	36
2.	UNIPI	CEA	9	36
3.	AUTH	CEA	9	36
4.	AUTH	UNISA	9	36
5.	UCL	UNIPI	9	36
6.	UCL	UNISA	9	36
7.	CEA	AUTH	9	36
8.	CEA	UCL	9	36
9.	UNISA	AUTH	9	36
10.	UNIPI	AUTH	9	36

11.	**AUTH**	**UNIPI**	**9**	**36**
12.	**UCL**	**UNISA**	**9**	**36**
13.	**CEA**	**UCL**	**9**	**36**
14.	**UCL**	**CEA**	**9**	**36**
15.	**UNISA**	**CEA**	**9**	**36**

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

	Main Training Events & Conferences	ECTS (*if any*)	Lead Institution	Project Month (*estimated*)
1	Establishing the foundations – SW Carpentry Bootcamp		SW Carpentry/CEA	M9
2	2 week school: "Data Science – an introductory school"	10	UNIPI	M15
3	2 week school: "Data Science in research: focus on particle physics and cosmology"	10	CERN/INAF/UCL	M18
4	Workshop on "Big Data: a wide view from industry projects"		AUTH	M30
5	2 days workshop "An introduction to Hadoop"		AUTH	M30
6	2 days workshop – Instructors course for SW Carpentry		SW Carpentry	M24
7	Conference on scientific harvest		UNISA	M41
8	Statistical or High Performance Computing School		UNISA	M20-M30

1.3 Quality of the supervision

Qualifications and supervision experience of supervisors

Students will be supervised for their research work by two supervisors, one in each university where they will obtain the PhD. In some cases mentors will provide additional expertise to complement the PhD supervision. All the supervisors from the beneficiary institutions have wide experience in PhD student supervision in their respective areas and have great scientific expertise to successfully guide the ESRs. The supervisors that will guide the ESRs are: Chiara Roda, Vincenzo Cavasinni, Francesco Fidecaro (UNIPI); Nikos Konstantinidis, Timothy Scanlon, Tomas Kitching (UCL), Rosy Nikolaidou, Jean-Luc Stark (CEA), Chara …..

Quality of the joint supervision arrangements (mandatory for EID and EJD)

The fellows will be based in two institutions designed as "Institution A" and "Institution B" and listed in table 1.2a. Two supervisors, one for each institution will guide the fellow with the topic and roles explained in section 3.1. The students will normally, but not necessarily, start the doctorate in "Instution A", that is also the recruiting institution, where the follow the doctorate courses and start the research project, they will spend a period in "Institution B" and return to "Institution A" to conclude the research project and prepare the thesis. The students will split their research time about equally between the two institutions of the double-degree. A third mentor fron a node where the ESR will be seconded will be assigned and together with the two supervisors will form the Supervision Committee (SC) that will monitor the student's progress. At the beginning of the program a personal file will be created to define scientific and training objectives and stages for their achievement. This will allow to effectively monitoring progresses. Meeting with supervisors will be frequent and at minimum once per week. Particularly attention will be given to promote gender balance in the mentoring relationships.

Non-academic contribution to the Supervision (mandatory for EID and EJD)

All ESRs 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.

X.1 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. Double degrees will be awarded by UNIPI, UCL, AUTH, CEA and UNISA. World leading researches in particle physics, astronomy/cosmology, gravitational waves research, statistical analysis, machine learning techniques and high performance computing from these institutes, will provide training via innovative research projects that will necessitate the development of groundbreaking techniques and foster new collaborations between all the nodes to fully exploit them. The training partners are world experts in their respective fields; for example, the Software Carpentry Foundation are pioneers in providing training on good software practices for scientists. The role of the industrial partners in the training of the ESRs is equally vital. In addition to mentoring 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.

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. The network builds, on the one hand, on the long-standing research collaborations between nodes and, on the other, enlarges and enriches existing collaborations bringing additional competencies. For example, in particle physics, the participating institutes have each expertise in the identification of different types of physics objects (leptons, jets, b-jets…) and putting this expertise together is a key ingredient for optimizing the sensitivity to the physics processes that will be studied. The interaction between the different research domains (particle physics, astronomy and gravitational waves research) will broaden the scientific horizons of the ESRs and has the potential to lead to breakthroughs in our understanding of nature. Finally, the interaction between academic and non-academic partners will foster the cross-fertilization of ideas, tools and techniques in data science employed in the two sectors, providing opportunities for good practices to be shared and adopted on both sides.

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, energy, 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

Since a few years, a debate is growing about the steady increase of the number of PhD in spite of the stable or decreasing pool of academic jobs[21]. Modernising the PhD path, by optimising the training to increase the employability of the PhD also outside academia, equipping them with soft skill, while preserving the mainstream goal of forming the new generation of scientists is mandatory. Big Data handling is an emerging field rapidly evolving under the needs of applied science, of business and economy and of fundamental research. Therefore a strong request for data scientists exists in many sectors. Europe appears to be behind USA, China, India and Brazil, in terms of the number of graduates with deep analytical training in 2008, according to a report of McKinsey Global Institute issued in 2011. This project aims at offering a modern training program that will keep pace with the rapidly evolving needs related to Big Data in science and society and to produce the human capital that will impact the competitiveness of Europe in science and will enhance the ability of European industry to pioneer development.

[21] J. Gould, "How to build a better PhD", Nature, vol 528, p22-25, 3 Dic 2015

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 more and more necessary to the research groups in the fields of particle physics, observational cosmology and gravitation. The ESRs enrolled in the BigDAPHNE program will acquire high-level training on Big Data, which they will apply to their respective field 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 domain.

Impact on non-academic career: The exposure to trends in Big Data handling and Big Data analysis tools, in the non-academic sector as in the research environment, along with a solid training in good code practices for the manipulation and handling of large data volumes exercised in the ESR research projects, will provide the young researchers with cutting-edge knowhow and will open entirely new perspectives to their career. The strong integration of the non-academic sector in the training program and the dedicated secondments at the companies' premises, will contribute to 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 Informatics Technology and application engineers, 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 companies' environment through their secondments, will allow them to discover opportunities for generating profit out of their knowledge and to develop entrepreneurial skills. Specialised training events are foreseen to invite the ESRs to actively consider them as main actors on the scene of enterprise[22]. In addition, through the strict time-lines that have to be respected by the experiments, the ESRs will develop project and time management skills. They will further 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.

Impact of the mobility actions: The mobility of the researchers is a built-in feature of this project. The BigDAPHNE network involves EU countries in very different economical and social situations: from the lively UK to Greece, which suffers a long-standing stagnation of economy. In addition, the academic institutions participating to BigDAPHNE have different histories; some of them are central poles of the national and EU academia, while others are peripherals young Universities building their reputation nowadays through the initiative of small active teams. 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

BigAPHNE provides a fantastic opportunity to build the necessary bridge between academia and industrial partners, which will be mutually beneficial. Several non-academic partners will contribute to training events and annual workshops, bringing the experience, the use cases of society and the solutions identified outside the research environment. The proposed training program involves secondments in the companies' sites which will provide the fellows interdisciplinary skills that may be imported in the research practice. A specific example is Semblent which has developed a visualisation methodology, strictly linked to the data structure and the data analysis techniques that the company implements, and might be considered for use in the research. Other valuable inputs are expected to come from Thales, NCC and other partners.

[22] HEP physicists, moved to making their own business in data analytics (http://squaresonblue.com), are available to discuss their experience with ESRs in such events.

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. The EJD program gives strong incentives to overcome these difficulties and enables to create lasting degree structures with critical mass that can compete for talented students worldwide. BigDAPHNE will aim at continuing the joint degree framework established by the project beyond its funding period. While the academic sector will seek funding from the well-known mechanisms national-wise, the involvement of the industrial sector will open new potential funding streams to complement this, thanks to the strong interest of the research-intensive industry that grows around Big Data analysis for the highly qualified professionals and the innovation of technology delivered by such training.

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

Dissemination of the research results

The standard international practice for the results of the scientific research 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 can be published in computational journals. The network will play also an active role in organising as a closing event a dedicated workshop on Big Data developments in physics research and outside academia, offering direct opportunities to the ESRs to present their achievements and the various stakeholders to convene and access the status of needs and new results in five years from now. All public results of the project, 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 and INSPIRE.

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 NCC, Software Carpentry, Semblent, through the secondment and trainings 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, ENEL, Thales, can further exploit software developments related to Big Data handling and Big Data analysis that will derive from the project. Already, through the secondment 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 (both civil servants and elected representatives), 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 include blogs regularly updated by the ESRs, short interviews and a regularly updated 'success storyboard' to emphasise research highlights and equal opportunities in research.
2. An emotional documentary, for Internet broadcast, will represent selected moments of the fashionable life and experience of the ESRs to increase of the attractiveness of scientific careers to

the young people. The film can also be presented in the scientific section of the international documentary film festival of Thessaloniki

3. The project will be promoted at science exhibitions, such as the Royal Society Summer Science Exhibition.
4. ESRs will promote their research at events such as the Researchers Night (funded by the EC) and Science Museum Lates. They will also be encouraged to attend events such as Pints of Science and Sceptics in the Pub, 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 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 the science. The ESRs will also participate in MasterClass events at their nodes.
2. Undergraduates: Meet a researcher style events will be hosted at the nodes to encourage more students to undertake research careers. These events will also focus on the ETN and how it helps improve their training and employability. Each ESR will undertake at least one such event.
3. Industry/Academia: We will hold a joint industry/academia workshop where 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 this week, which has a mixed audience of industry and academics from a range of fields.
4. Policy makers: Attend events focused on promoting science and its benefits to civil servants and elected representatives, such as the Parliamentary Links 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 four research/training WPs and three management WP  The first two WPs focus on research projects in **"Particle physics" (WP1)** and **"Cosmology and Gravitation" (WP2)**. The leading beneficiary of WP1 is CEA institute, represented by Dr. R. Nikolaidou. She is a senior member of ATLAS, a world-expert in Higgs physics and a previous PI of the ARTEMIS-ITN project in FP7. The already existing ties, through past collaboration, between the ATLAS members of all the beneficiary institutes of the network, provides to Dr.. Nikolaidou a strong, competent and complementary network of experts to efficiently support the ESRs in their cutting-edge research projects. UCL, the WP2 leading beneficiary, coordinates the research projects on deep sky surveys and on gravitational waves. Dr. T.Kitching presents an optimal profile to coordinate WP2, as he currently covers leading roles in the scientific organization of EUCLID.. Members of the other institutions (CEA, UNIPI, INAF, AUTH) in the network, involved in WP2, are key components of the EUCLID and VIRGO collaborations and provide complementary expertise that will ensure a world-leading support for the ESRs research projects in cosmology. Scientists of UNISA are in the process of joining the EUCLID consortium bringing their previous expertise gained in previous related projects,

WP3 and WP4 deal with the technological part of the research projects and address **"Big Data managing" (WP3)** and **"Big Data analysis" (WP4)** respectively. WP3 is led by AUTH, chaired by Prof. C. Petridou. She is a senior physicist with extensive experience and expertise on coordination in many experiments. She will exploit and profit from the strong technical support provided by the Scientific Computing section of the IT center of AUTH. Dott. I. Salmatzidis, who is the Technical Manager of the Information Technology Centre of AUTH and dott. P. Korosoglou, who is the Technical Coordinator of the Scientific Computing Office of AUTH, with their well-recognized experience on Big Data training around PRACE and the first Greek Supercomputer ARIS, will guarantee a state-of-the-art training on Big Data managing for all ESRs. WP4 is led by the UNIPI where Prof. D.Pedreschi, a world-expert and pioneer researcher in Data Science, brings the competence and expertise to lead and coordinate this WP. Additional expertise to WP3 and WP4 will be added by CERN, CNR, INFN and UCL also having strong background in IT technologies and training on data analytics.

The research projects proposed for the ESRs of BigDAPHNE are all designed to address the most pressing questions in the scientific landscape of fundamental physics today: the origin of dark matter and

dark energy, the experimental evidence of gravitational waves and the physics beyond the Standard Model. All ESR projects are connected to WP1 (ESR1 to ESR9) or to WP2 (ESR10 to ESR15). Connections between ESRs are described in the individual project description in Tables 3.1d.
WP3 and WP4, dealing with the technological challenges related to Big Data, involve all ESRs in order to (1) encourage them to critically review the frameworks, tools and methods used in their work and (2) to equip them with competent cutting-edge technical knowledge to assess alternative solutions and possible advancements in Big Data issues. A key ingredient in this path to build Big Data Scientists, is the secondment experience, where the ESRs will be asked to work on a Big Data Project technically linked to the same challenges they face in their research project, but, at the same time addressing completely different questions.
The three additional WPs deal with **"Training" (WP5)**, **"Dissemination and Outreach" (WP6)** and **"Coordination and Management" (WP7)**. WP5 is designed to oversee all the training activities in the network (schools, workshops, seminars, conferences) and is coordinated by UNISA, led by Prof. S. Spagnolo. She has great expertise in coordination (representing the INFN-Lecce and UNISA HEP experimental group in the Scientific Committee 1 of INFN in the past four years and co-organizing the INFN Summer School of Otranto in the recent years) having therefore a very good profile to efficiently lead this WP. She will also rely on the wide and diverse training experience of the many members of the network documented in Tables 4.1. WP6 focuses on Dissemination and Outreach. 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 Science and IT technologies. UNIPI is leading WP6 with Prof. V. Cavasinni providing the perfect curricula to organize this WP, since he has vast experience in the organization of scientific outreach events. Moreover many members of the network are familiar with delivering seminars in high schools or public venues or to participate to the "master-class" or "researcher's night" events, providing thus a large pool of people to rely on, for the organization of outreach and dissemination events. The SW Carpentry path that all ESRs will undertake gives particular emphasis to the dissemination part of the project.
Finally, WP7 deals with "Coordination and Management" and is leaded by Prof. C. Roda (UNIPI). She has large experience on handling budgets of the order of 500k euro/year (salary excluded) and management skills in leading and organizing groups both at national and international level.

Table 3.1 Description of Work Packages

WP Number	1	**Start Month – End Month** 3-48
WP Title	Frontier research in particle physics	
Lead Beneficiary	CEA	
Objectives: Use full LHC data-set delivered until end of 2018 to: (1) search for Dark Matter signatures at hadron colliders; (2) measure Higgs boson couplings with high precision (3) search for new physics in diboson topologies. All these studies will be carried out with advanced data handling and analysis techniques.		
Description of Work and Role of Beneficiaries / Partner Organisations A number of analyses will be carried out to search for new physics in a wide range of experimental signatures. The analyses will look for deviations from the Standard Model predictions in diboson production or in Higgs boson couplings. The analyses can be organized in three categories: single Higgs boson events (H->4l, H->bb), Higgs pairs (4b and 2tau2b) and dibosons (WZ, WW, ZZ…). 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 full 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 cross-checks for possible new physics. **Task1.1** Optimization of the selection of physics objects (leptons, jets, taus…) and of tools for systematic uncertainty definition. Institutions contributing with complementary expertise: leptons (AUTH, UNISA, 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) with details on the interactions between the two ESR projects for cross-checks and exchange of expertise on the many common items (ESR1, ESR5). **Task1.3** Higgs coupling: definition of the analysis strategy for the associated VH(bb) production (lead UCL ESR6), H->4l (lead CEA) and ttbarH(bb) (lead CEA) with details on the interactions between the ESR projects for cross-checks and exchange of expertise on the		

many common items (ESR6, ESR7, ESR8).
Task1.4 Diboson searches: definition of the analysis strategy for the diboson production searches in the leptonic (lead AUTH), leptonic and missing ET (lead AUTH), semi-leptonic (lead UNISA) and all hadronic (lead UNIPI) final states. Details will be defined on the overlap and cross-checks between the projects (ESR2, ESR3, ESR9, ESR4).
Task1.5 Review of type and size of simulated datasets to be produced for the various analyses. ESR1-9 involved.
Task1.6 Survey of new tools for event selection and statistical analyses and assessment of improvements in sensitivity. ESR1-9 involved.

Description of Deliverables
1.1 Report on the analysis strategy and ESR interactions for the three types of searches (Double Higgs production, Higgs coupling measurements, exotic diboson searches) - M15
1.2 Report on the novel approaches for analysis selection and statistical analysis for the three types of searches (Higgs pair production, Higgs coupling measurements, exotic diboson searches) – M24
1.3 Measurement of double Higgs production or derivation of cross-section limits in HH->4b, HH->2tau2b and their combination - M48
1.4 High precision measurement of HZZ coupling – M48
1.5 Measurement or limit assessment of the ttH coupling – M48
1.6 Results on new physics searches in diboson final states – M48

WP Number	2	**Start Month – End Month** 3-48
WP Title	Frontier research in Cosmology and Gravity	
Lead Beneficiary	UCL	

Objectives: Perform a large number of HPC simulations of binary neutron star mergers and develop efficient techniques to analyze a large volume of gravitational-wave data with a matched-filtering technique. Establish on simulation methodologies for an efficient implementation of the analysis of 3D weak lensing data that EUCLID will deliver. Define statistical procedures for the simultaneous extraction from data of large numbers of model parameters and nuisance parameters related to systematic uncertainties. Combine EUCLID dark matter sensitive data with other catalogues using novel methodologies for an efficient merging of information.

Description of Work and Role of Beneficiaries / Partner Organisations 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.
In the EUCLID UCL and CEA teams a wide expertise exist on 3D weak lensing. Experience with Bayesian unfolding, 2D lensing, at the strong and weak scale, is wide spread in the groups. Based on the existing knowhow, 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.
Task2.? 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.
Task2.? 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). ESR11 involved.
Task2.?
Task2.?

Description of Deliverables
2.1 Report on the implementation of equations of state in the 3D code
2.2 Report on the extraction of damping timescales through numerical simulations
2.3 Report on the implementation of analytic gravitational-wave templates
2.4 Report on the applying the templates on real data searches

WP Number	3	**Start Month – End Month** 3-48
WP Title	Big Data managing in industry and research	
Lead Beneficiary	AUTH	

Objectives: Based on the research projects and secondment projects (1) acquire a critical view on the tools and framework used for data handling (storing, processing, monitoring, managing movement), (2) get trained to learn trends and new approaches from different environments; (3) identify strategy and ESR projects suitable for optimization and (4) implement optimization and assess performance

Description of Work and Role of Beneficiaries / Partner Organisations
The fellow will use their research and secondment projects as benchmarks to acquire a critical view on the way the datasets they use and produce (after reconstruction, after pre-processing, …) are managed. They will analyse the available tools and investigate the possible use of these tools to improve the existing frameworks. After this survey ESR projects where key management phases can be optimized will be identified and we will proceed to implementation. The performance improvements obtained will be assessed. Focused view and exchange of ideas, use of approaches and tools from different environments are the key factors for this WP.
Task3.1 Investigate the structure and tools for the data managing steps (production, processing, movement, management) in ATLAS (ESR1-9), Virgo (ESR10, ESR11) and EUCLID (ESR12-15) (lead AUTH and in collaboration with UCL, UNIPI, CEA, UNISA).
Task3.2 Investigate the use of parallelized data processing, role of High Performance Computing (ESR10-15) (lead AUTH and in collaboration with UNISA, INAF, UCL).

Task3.4 Identification of pool of ESR projects to proceed with data management optimization.
Task3.5 Optimization implementation and performance assessment.
Task3.6 Evaluate and report on exporting of tools and methods from research environment to secondment projects and vice versa (lead AUTH in collaboration with partners).

Description of Deliverables
3.1 Report on the structure of the data managing in the research projects (All ESR) and identify critical points if any – M12
3.2 Reports on possible solutions to improve data managing (parallelization, monitoring of processing, data distribution and storing …) and identification of ESR projects to proceed with optimization – M24
3.3 Report on implementation and on performance achievement in speed, reliability, scalability of research project frameworks – M36
3.4 Report on exporting of tools and methods from research environment to secondment projects and vice versa – M40

WP Number	4	**Start Month – End Month** 3-48
WP Title	Big Data analysis in industry and research	
Lead Beneficiary	UNIPI	

Objectives: Based on the research projects and secondment projects (1) acquire a critical view on the tools and methods used for data analysis; (2) get trained to learn trends and new approaches on data analysis from different environments; (3) identify strategy and ESR projects suitable for optimization and (4) implement optimization and assess performance.

Description of Work and Role of Beneficiaries / Partner Organisations
The fellows will use their research and secondment projects as benchmarks to acquire a critical view on the way the information is extracted from the data in different environments. They will first focus on a survey of the characteristics and performance of the available methods and tools based both on the information acquired during the training and on the existing framework in the various research environments. The survey will cover aspects as high-rate event selection algorithms for rare events, advanced data techniques statistical methods, multivariate analysis, machine learning methods, use of high performance computing… They will than decide which one(s) of these tools would improve the existing frameworks and define which ESR projects are suitable for this optimization. For these projects we will proceed to optimization and performance assessment. Focused view and exchange of ideas, approaches and tools from different environments are the key factors for this WP. Particularly interesting will be the collaboration between cosmology and particle physics ESRs.
Task4.1 Survey on the structure and tools used for the data analysis (online and offline) in ATLAS (ESR1-9), Virgo (ESR10, ESR11) and EUCLID (ESR12-15) (lead UNIPI in collaboration with UCL, AUTH, CEA, UNISA)
Task4.2 Feasibility study to improve the analysis sensitivity with new analysis methods: Machine learning methods (ESR1-9), 3D-spin wavelet (ESR14), Bayesian hierarchical methods (ESR12)…
Task4.3 Identification of pool of ESR projects to proceed with data analysis optimization.
Task4.4 Optimization implementation and performance assessment.
Task4.5 Evaluate and report on exporting of tools and methods from research environment to secondment projects and vice versa (lead AUTH in collaboration with partners).

Description of Deliverables
4.1 Report on the structure of the data analysis and used tools in the research projects (All ESR) – M12
4.2 Reports on possible solutions to optimize data analysis and identification of ESR projects to proceed with optimization – M24
4.3 Report on implementation and on performance achievement for the ESR projects identified at point 4.2 – M36
4.4 Report on exporting of tools and methods from research environment to secondment projects and vice versa – M40

WP Number	5	**Start Month – End Month** 1-48
WP Title	Training	
Lead Beneficiary	UNISA	

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) help identifing required additional training.

Description of Work and Role of Beneficiaries / Partner Organisations
ESR will participate many events organized by the network as described in Section 1.2. All these events need to be organized timely to fulfill 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 overseen by this WP in order to be sure they are planned at the best time.
Task 5.1 Coordination of the training events. WP5 will be responsible for the timely organizations of the workshops. The AB will setup a program committee for each event of the network. The committee will consist of WP5 leader, experts of the network and a local scientific responsabile. This committee will be responsible of the scientific program and of the timely delivery of the workshop. A local committee will take care of logistic aspects.
Task 5.2 Setup an event evaluation procedure and collect ESR feedback to be distributed to the SB and PO.
Task 5.3 Oversight secondments: an initial indicative secondment plan has been prepared. Once the ESR are recruited and the SC are in place the second plan will be revised. WP5 leader will collect all proposal for revision and communicate the revised plan to AB and PC.
Task 5.4 Identify courses for elective schools and in case of request additional training. WP5 leader will keep AB informed of the definition of elective courses and of additional requests.
Task 5.5 Ensure publication of training material on the network website.

Description of Deliverables
5.1 Report on the BigData schools with feedback from the ESRs – M30
5.2 Report on effectiveness of secondment plan – M40

5.3 Report on SW full cycle effectiveness with feedback from ESRs – M40
5.4 Training material available on website – M48

WP Number	6	**Start Month – End Month** 1-48
WP Title	Dissemination and Outreach	
Lead Beneficiary	UNIPI	

Objectives: (1) Scientific and technological context communication to the public; (2) educate fellows to communicate science and technique with effectiveness; (3) attract public to science; (4) disseminate the scientific results to different scientific communities.

Description of Work and Role of Beneficiaries / Partner Organisations
This WP has a twfold aim of communicating to the public the scientific and technological challenges and to present the scientific results to a large scientific community. WP6 leader is responsible for a carefull planning of the dissemination and outreach events that are described in Section 2.4. WP6 leader is also responsible of a timely deliver 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 program. Particular attention will be put in higlighting the experience with the ESR work in both academic and non-academic environments.
Task 6.1 Provide website content and assist PO in drafting recruitment material.
Task 6.2 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 6.3 Advertise the program and the planned dissemination and outreach events.
Task 6.4 Dissemination of results in the scientific community at large and in particular in the two network conferences.

Description of Deliverables
6.1 Dissemination and outreach activity plan – M4
6.2 Annual reports on dissemination and outreach activities at annual meeting – (M12, M24, M36, M48)
6.3 First short documentary on the project – M15
6.4 Scientific publication report from WP1-WP2 – M48
6.5 Dissemination and outreach material on Website - M48
6.6 Scientific conference report – M48

WP Number	7	**Start Month – End Month** 1-48
WP Title	Managing 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 maintaing comunication with EU on the activity and finatial status of the project.

Description of Work and Role of Beneficiaries / Partner Organisations
A Program Office (PO) will be setup at UNIPI an 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. A sub-office of the PO, the Graduate Office (GO) is designed to take care of all issues related to EJD (recruitment, quality control,…).. The main tasks of this WP are:
Task 7.1 Coordinate the installation of all the Boards and committees of the project in a timely manner;
Task 7.2 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 7.3 Make sure tha 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 7.4 Support WP7 and WP8 leaders in the organization of training, diseemination and outreach;
Task 7.5 Coordinate with doctorate offices at double-degree awarding University offices.

Description of Deliverables
7.1 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 – M0
7.2 Kick-of meeting at UNIPI – M1
7.3 Recruitment report with information on recruited ESR – M5
7.4 European Joint Degree Agreement (EJDA): Signed detailed agreement between each two universties awarding the double-degree specifying rules to obtain double PhD degree in the program – M3
7.5 Declaration of conformity for each ESR – to be clarified
7.6 Career Development Plan for ESR prepared – M
7.7 Report on PhD awarded degrees with short description of the thesis – M48
7.8 Final project report with all activities and results obtained – 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	WP3,WP1,WP2	AUTH	M24	Produce a document with list of ESR

	for Data Management framework optimization				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	Some milestone for Virgo	WP2			
M2.2	Some milestone for Virgo	WP2			
M2.3	Some milestone for EUCLID	WP2			
M2.4	Some mileston for EUCLID	WP2			
M5.1	Big Data schools completed	WP5,WP3,WP4	UNISA	M25	Two schools completed with material uploaded to the website
M5.2	SW carpentry path completed	WP5,WP6	UNISA	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	UNISA	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 BigData 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	Thesis submission	WP7	UNIPI	????	

Table 3.1 d Individual Research Projects

Fellow *ESR1*	Host institution UNIPI	PhD enrolment UCL/UNIPI	Start date *Month 9*	Duration *36 months*	Deliverables *(Refer to table 3.1b)*

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 (e.g. Phys. Rev. D92, 092004 (2015)). With the new LHC run, at higher energy and luminosity, the HH sensitivity, within the standard model physics, as well as the discovery potential for new physics will be greatly increased. In particular the HH->2b2τ topology, together with the HH->4b decays (ESR5), has been already shown to be 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. The optimal identification of b-jets is also instrumental for the sensitivity of this channel and this work will benefit from the synergies and complimentarities with the work of ESR5 and ESR6.

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):

- *NCC: 4 months* – Network activity data management and online analysis

Fellow *ESR2*	Host institution UNIPI	PhD enrolment UNIPI/CEA	Start date *Month 9*	Duration *36 months*	Deliverables *(Refer to table 3.1b)*

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: After the Higgs boson discovery and thanks to the highest proton-proton collision energy ever reached, the search for new physics has become the main focus at the multipurpose LHC experiments (ATLAS and CMS). The fellow will enter this very challenging community to play a leading role in the search of new physics in events with missing transverse energy and a hadronically decayed heavy boson. This signature takes advantage of the high signal rates thanks to the high branching ratio but has to cope with the high multi-jet and boson+jet background. This kind of 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 Triple, Gravitons, W'…). In both cases, the hadronically decaying boson is searched for in the boosted regime where the couple of quarks coalesce in a single "FatJet" at detector level. 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. One of the key issues for these signatures is the trigger selection. 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 a challenge 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 allow to either discovering of new physics or setting of world best limit on many new physics models. This work has synergies with and complements ESRs 1, 4, 5 and 6.

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 institution AUTH	PhD enrolment AUTH/CEA	Start date *Month 9*	Duration *36 months*	Deliverables *(Refer to table 3.1b)*

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 institution AUTH	PhD enrolment AUTH/UNISA	Start date *Month 9*	Duration *36 months*	Deliverables *(Refer to table 3.1b)*

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. The latter being a beyond the SM expansion of the Lagrangian of the system in terms of dimension -6 or -8 operators, which need to be measured experimentally. 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 necessary to be developed, in order to extract results by properly handling the correlation between the uncertainties of the various channels. 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):

- AUTH1: 4 months – Intrusion detection and alerting service

Fellow *ESR5*	Host institution UCL	PhD enrolment UCL/UNIPI	Start date *Month 9*	Duration *36 months*	Deliverables *(Refer to table 3.1b)*

Project Title and WP(s) to which it is related: New physics in HH->4b with ATLAS (WP1,WP4)

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):

- Semblant: 4 months – Data analysis with unstructured or semi-structred data

Fellow *ESR6*	**Host institution** UCL	**PhD enrolment** UCL/UNISA	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(Refer to table 3.1b)*

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, 2, 5 and 7.

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):

- Transport for London (TfL): 4 months – Optimization of London Undergourn Asset Maintenance Strategy

Fellow *ESR7*	**Host institution** CEA	**PhD enrolment** CEA/AUTH	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(Refer to table 3.1b)*

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 program 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):

- UBIS: 4 months – Data Analysis for banks (to be better defined)

Fellow *ESR8*	**Host institution** CEA	**PhD enrolment** CEA/UCL	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(Refer to table 3.1b)*

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 institution** UNISA	**PhD enrolment** UNISA/AUTH	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(Refer to table 3.1b)*

Project Title and WP(s) to which it is related: New phenomena in multi boson final states at ATLAS (WP1, WP3, WP4)

Objectives: Multi-boson final states are an interesting scenario to look for new phenomena: several theory-motivated models of physics beyond the SM assume new symmetries, which might lead to resonant production of pairs of gauge bosons. Additional Higgs-like particles might also lead to deviations of the data from the SM expectation for multi boson production. In addition the production of two bosons plus two forward jets open a window on vector boson scattering. While fully leptonic decays of gauge boson lead to clean signatures, their sensitivity is statistically limited in comparison with the hadronic or semi-leptonic signatures. This research program will investigate the semi-leptonic final states using the modern techniques of vector boson tagging via the reconstruction of a fat-jet (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 fat-jet, 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 to provide discrimination against background for vector boson identification based on the fat-jet features and for heavy-flavour tagging applied to these final states. Therefore novel methodologies can be fruitfully applied to this analysis. The work will benefit from the synergies and complementarities with the work of ESR4, ESR5 and ESR6.

Expected Results: (1) Implement of modern machine learning techniques for optimal signal-to-background discrimination. (2) Publish ATLAS results on VV production with interpretation in various theoretical models. (3) Re-assess sensitivity to anomalous Quartic Gauge Couplings at HL-LHC.

Planned secondment(s):

1. ENEL: 4 months – Energy Management for trading department

Fellow ***ESR10***	**Host institution** **UNIPI**	**PhD enrolment** **UNIPI/AUTH**	**Start date** ***Month 9***	**Duration** ***36 months***	**Deliverables** ***(refer to numbers in table 3.1b)***

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 templates, coming from numerical simulations and/or analytical models. Post-merger parameter estimation will be investigated using real data coming from the advanced LIGO-Virgo network. The main goal will be to develop efficient techniques to analyze a large volume of data with a matched-filtering technique. Observational constraints on the radius of neutron stars will be obtain and depending on the outcome the possibility for constraining alternative theories of gravity will be investigated.

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): ESR10 is expected to perform most of the training and research described above at the partner organization INFN Pisa, supervised by Giancarlo Cella. Three 3-month long stays at AUTH under the supervision of N. Stergioulas will be required for co-supervision on the physically-motivated templates, as well as two 1-month stays at HITS (A. Bauswein) for the same purpose. Furthermore, a three-month stay at the partner-organization Georgia-Tech (J. Clark) will be useful for in-depth training on required numerical techniques.

- CERN 4 Months – Data Project withing OpenLab to be better defined

Fellow ***ESR11***	**Host institution** **AUTH**	**PhD enrolment** **AUTH/UNIPI**	**Start date** ***Month 9***	**Duration** ***36 months***	**Deliverables** ***(refer to numbers in table 3.1b)***

Project Title and Work Package(s) to which it is related: High-Performance-Computing simulations of binary neutron star mergers

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. ESR11 will implement a new, efficient method for handling tabulated, hot equations of state in the Einstein Toolkit. The simulations will be performed on the PRACE network. The main goal will be to accurately extract the damping timescale of gravitational waves as well as of quasi-radial oscillations for the first 10 milliseconds after merger. A large number of simulations with three different resolutions will be performed. Correlations between damping timescales and 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): Three 3-month long stays at UNIPI/INFN Pisa under the co-supervision of Giancarlo Cella will be required for transferring the results of the simulations to the data analysis effort. Two 1-month stays at HITS (A. Bauswein) will be required for training on hot equations of state. A three-month stay at the partner-organization Georgia-Tech (J. Clark) will be useful for in-depth training on required numerical techniques.

- ENEL 4 Months - Smart City Malaga - Consumption profile

Fellow *ESR12*	**Host institution** UCL	**PhD enrolment** UCL/UNISA	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(refer to numbers in table 3.1b)*

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 build 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 Bayesian Hierarchical 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. Essentially this project will create a probabilistic model of the galaxy, telescope, instrument and measurement chain. 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 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 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): *UCL, supervisor : T. D Kitching, J. McEwen, Partner (INAF, Rome) Francesco Strafella and Roberto Scaramella.*

- Semblant 4 Months – Tools for Spark and Cassandra – to be better defined

Fellow *ESR13*	**Host institution** CEA	**PhD enrolment** CEA/UCL	**Start date** *Month 9*	**Duration** *36 months*	**Deliverables** *(refer to numbers in table 3.1b)*

Project Title and Work Package(s) to which it is related: 3D Dark Matter Mapping

Objectives: TBW

Expected Results: TBW

Planned secondment(s): *CEA, supervisor: J-L. Stark, Partner (UCL): T. D Kitching, J. McEwen*

Fellow *ESR14*	Host institution UCL	PhD enrolment UCL/CEA	Start date *Month 9*	Duration *36 months*	Deliverables *(refer to numbers in table 3.1b)*

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 therefore traditionally requires a spherical-Bessel transform. However this is inefficient and sub-optimal in that case that there are complex masks in the data, and that there are 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 Euclid dark energy measurements, and extend the approach to account galaxy alignment and instrumental systematic effects (a further extension would be to generalise this to include Cosmic Microwave Background cross correlations). 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 systemati effects in the data, and then to measure the dark energy equation of state as a function of redshift.

Planned secondment(s): UCL, supervisor: T. D Kitching, J. McEwen, *Partner (CEA, Paris): 6 months*

- CERN 4 Months – Parallel programming to be better defined

Fellow ESR15	Host institution UNISA	PhD enrolment UNISA/CEA	Start date *Month 9*	Duration *36 months*	Deliverables *(refer to numbers in table 3.1b)*

Project Title and Work Package(s) to which it is related: Combining the future multicolor 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 modeling 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) Modeling 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): CEA, supervisor: J.L. Stark, Partner (INAF, Rome): 6 months (purpose....)

- AUTH 4 Months: Environmental monitoring and energy efficiency

3.2 Appropriateness of the management structures and procedures, including quality management and risk management (with a mandatory joint governing structure for EID and EJD

projects)

- · Supervisory board

The **Supervisory Board (SB)** is the governing body of the programme, is responsible for delivery of all aspects of the project and is the ultimate arbitrator. It will be made up of representatives of all beneficiaries and partner organisations. The ESRs will also elect one of their number to sit on the board. The composition of the SB will ensure that a full range of views are heard and taken into account when any decisions are made. The SB will be chaired by the Program Co-ordinator (PC), with decisions taken by a simple majority. In the event of a tie the PC will have the deciding vote. The SB will meet once for the kick off meeting, then annual, with the option to hold extraordinary meetings in exceptional circumstances.

The SB will monitor the network and ensure the project objections are being achieved. The SB will oversee all network wide activities: training, financial, administrative, outreach, dissemination, integrity of research, progress of ESRs, industrial placements, IPR and ensure the budget is distributed as set out in the CA. The SB will also be the ultimate arbitrator in the unlikely event of any disrupts or conflicts. The SB will also monitor and decide upon areas where the programme could be improved.

- · Network organisation and management structure, including financial management strategy, strategy for dealing with scientific misconduct

The **Program Office (PO)** will act as the project management office, which will implement all decisions reached by the SB and will be responsible for the day-to-day management of the programme. The PO will be responsible for the managing the operational finances, preparing all administrative and financial reports, for both internal and external consumption. It will issue quarterly financial reports to the members of the SB, present an annual review at SB meetings and will be responsible for all communications with the EC. The PO will also oversee the management of the intellectual property rights (IPR). The PO will be hosted at the PC's institution, where suitable structures already exist. According to the needs of the PO, it will be staffed with an extra officer, which will be jointly funded by the beneficiary institutes.

The **Graduate Office (GO)** will be a sub-office of the PO and will assist the AB in overseeing the EJD. It will manage the recruitment process, provide quality control of the Ph.D. programmes, maintain records on the ESRs and manage communication between the degree awarding bodies.

Each work package (WP) will have one **Work Package Leader (WPL)**. The WPL will be responsible for co-ordinating the activities of that WP and are tasked with ensuring all milestones and deliverables are achieved in accordance with the project plan. They will organise WP specific meetings, ensure all participants are working in a co-ordinated manner, monitor progress, liaise with the PO and prepare annual reports on the WP progress, which will be ratified by the SB.

Scientific integrity is of the upmost importance to this programme. If any misconduct is identified it will be reported to the AB. The AB will hear the case and pass judgement. Depending on the seriousness of the case, possible sanctions will range from censure to expulsion from the programme. If the accused feels they have been unfairly treated they can appeal to the SB who will be the ultimate arbitrator of any disputes and will ratify any decisions made by the AB.

[Figure will be included here showing links between ESRs/PO/SB]

- · Joint governing structure (mandatory for EID and EJD projects)

An **Academic Board (AB)** will be responsible for overseeing the Ph.D. programme, ensuring the EJDA is fully implemented and academic standards are maintained. It will be composed of one member from each degree granting institute. The work of the AB will be supported in an administrative capacity by the GO and will be monitored by the SB.

The AB will form several committees in order to effectively execute its remit:

- **Admissions Committee (AC)**: Responsible for setting the admissions criteria and producing a ranked list of candidates for each of the ESRs. It will be formed of the AB plus additional facility members and suitable personnel from the industrial partner companies chosen for their particular expertise.

- **Supervisory Committee (SC)**: Each ESR will have a supervisory committee formed of their supervisors and a representative from the institution hosting their internship. The SC will be responsible for monitoring, evaluating, supervising and mentoring the ESR. The SC will be the first body to resolve any issues for the ESR. In the event that the SC is unable to resolve the issue, it can 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 their representative on the board, which provides an independent channel of communication. The SC will be responsible for ensuring that all institutional requirements are meet.

- **Thesis Committee (TC)**: A TC will be formed for each ESR and is tasked with examining the thesis according to the EJDA rules. The five member TC will be formed of the SC, one qualifying facility member from the degree awarding institutes and an external academic examiner.

- · For EJD, joint admission, selection, supervision, monitoring and assessment procedures

Joint admission

The AC will set the joint admissions and selection procedure and ensure it is unified across the network. All posts require a strong background in physics and a Masters degree from an accredited institute. A strong background in experimental or computational skill will be advantageous, but will not be required. A common set of supporting documents will be obligatory: curriculum vitae (maximum length 2 pages), a statement of research interest (maximum 1 page), two letters of reference and a transcript of their grades. As English will be the language of the network all applications and supporting documentation will be required in English. In certain circumstances applicants maybe required to provide an internationally recognised English language diploma certificate to ensure that they will be able to effectively participate in the network.

Additional complications arise from the need for all applicants to pass the requirements of both the degree awarding institutes. All applications will first be made to the network, where they will be processed by the GO. The AB, with input from the AC, will then certify that the minimum requirements of the programme and both the degree awarding institutes have been meet. The application will then be forwarded onto the two relevant institutes.

The posts will all be advertised shortly after the project has been accepted, but before the project has officially begun (aimed for February 2017), to ensure there is ample time to advertise and fill all the posts. The aim is to have all posts filled by April 2017 (M3) to begin in September 2017 (M9). This should provide sufficient redundancy in case of difficulties filling all the posts. To facilitate the early advertisement of the posts, the programme website and online application tool will be available by October 2016.

Selection

The AB will produce a shortlist of candidates for each ESR based on their credentials. The short listed candidates will then be interviewed by a selected subset of the AC with expertise relevant to the ESR. The full AC will meet to discuss the shortlisted applicants and rank them based on the interview outcome and their credentials. The rankings will be ratified by the SB before the top ranked candidate are notified of the outcome. If the top ranked candidate declines the position, then the next highest ranked candidate will be approached.

Candidate selection will obey all recruitment laws and will adhere to the code of conduct for the recruitment of researchers, as outlined in the European Charter for Researchers.

Supervision, monitoring and assessment

Each ESR will have a SC assigned to them by the AB who will regularly monitor and assess their progress, as outlined in the section on 'Joint governing structures'. The SC will have a representative from both degree awarding institutes and the partner company. The ESR's progress will also be overseen on a day-to-day by appropriate members of the SC depending on the stage of the project. The SC will produce more formal quarterly reviews for the AB and an annual review for the SB. The AB will meet quarterly and the SB annual to review these progress reports.

- · Recruitment strategy

Emphasis will be put on recruiting the best possible candidates regardless of origin. Procedures will be put in place to ensure that all hiring is carried out in an unbiased manner adopting best equal opportunity practises and adhering to all relevant laws. The positions will be advertised aligned with the academic year to maximise applications. A wide-ranging advertisement campaign will be carried out encompassing both a targeted and a more wide-ranging approach:
- Specially produced posters, announcements and e-mails targeting the beneficiary institutes and other top research institutes worldwide.
- Use of websites such as the EU portal (EURAXESS), jobs.ac.uk and inspirehep, and those associated to journals (Physics Today, Nature, Science and New Scientist).
- Advertise positions on social media accounts (for instance Facebook and Twitter) of both the partner and beneficiary organisations.
- Make full use of national and international research contacts and networks to propagate advertisements.
All advertising will highlight the unique nature of the ESRs, the high calibre of companies at which internships will be carried out, the excellence of the research and the highly increased employability of graduates from this scheme.

- · Progress monitoring and evaluation of individual projects

The ESRs will be monitored and evaluated by their SC. The supervisors will meet regularly (it is envisaged that this will be at least weekly) with their ESRs on a one-to-one basis to discuss their project. More formal monitoring and evaluation will also be carried out by the SC in quarterly reviews based on the ESR's CDP. This will evaluate the progress along the CDP, provide updated milestones, ensure all institutional requirements have been meet and will allow the ESR to highlight any issues they would like to raise. A short report jointly signed by the ESR and SC will be produced. A copy of these reports will be provided to the GO. An annual review will also be prepared for each ESR and presented to the SB at the annual meeting.

- · Risk management at consortium level (including table 3.2a)

Four main areas of risk have been identified:
R1 Delay in recruitment (risk: low): The recruitment of the ESRs has been aligned with the annual recruitment cycle for Ph.D. students. Given the recruitment strategy outlined above, the fact that all the degree awarding partners have a high Ph.D. application rate for these subject areas and the prestigious nature of these posts, it is deemed highly unlikely that they will not be filled. In the unlikely event that all posts are not filled, subsequent hiring rounds will be held. If it is not possible for ESRs hired in subsequent rounds to attend the earlier training courses, then they will be provided with the online material and will be sent to similar courses held at other nodes in the network (XXX add specifics here).
R2 ESRs unable to complete project (risk: low) – Only ESRs who have a very high academic standard will be recruited. They will be undertaking a research project designed by highly experienced researchers who have supervised many Ph.D. students. They will also be embedded into a research group where more advanced graduate students, as well as postdoctoral researchers, who will be available to provide help and advice. If an ESR is however unable to complete his project, a replacement at a similar stage of research will be recruited from within the network.
R3 ESRs unable to complete Ph.D. during the duration of the project (risk: medium) – Despite the high level of supervisions and monitoring, due to the nature of research, there is a possibility that a project may take longer than envisaged. In the case that an ESR will not complete within three years all participants have committed to the programme till all ESRs have finished their degrees regardless of the

end date of the project.

[TS - We could add a few more here, perhaps (1) partner company dropping out (2) issue with data collection at experiments]

- · 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 adhere to 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 meeting to promote IPRs to all the academic participants of the programme, to encourage the commercial exploitation of the research.

- · Gender aspects

Gender equality and the promotional of women in science are given a very high priority in this ETN. In high energy physics and cosmology women only generally only make up ~20% of faculty positions. The composition of the senior positions in this proposal significantly surpass this number. The PC and four out of five of the institutional representatives will be female. The AB and SB will also be composed of XXX% women. The consortium is committed to attract female fellows and will exploit contacts to ensure we receive as large as possible number of applicants from women. Profiles of the leading female researchers and fellows will be highlighted on the networks webpage to promote greater involvement of women in these fields which are still largely male dominated environments.

- · Data management plan (*only if participating in Open Research Data pilot – see page 18 above*) [I assumed this doesn't apply to us…]

Table 3.2a Implementation Risks

Risk No.	Description of Risk	WP Number	Proposed mitigation measures
R1	Delay in recruitment	WP1-4	- Advertisement of posts begins before programme officially starts. - Additional recruitment rounds held will be held till all posts filled. - Possibility for some ESRs to start project later.
R2	ESRs unable to complete project	WP1-4	- Strong supervisory structure in place. - Replacement candidates will be recruited from within the network.
R3	ESRs unable to complete Ph.D. during the duration of the project	WP1-4	- Strong supervisory structure in place to identify and remedy any such issues at an early date. - Commitment from all participants to continue beyond end of programme until all ESRs have completed their degrees.

The following sections of the European Code of Conduct for the Recruitment of Researchers refer specifically to recruitment and selection:

Recruitment

Employers and/or funders should establish recruitment procedures which are open, efficient, transparent, supportive and internationally comparable, as well as tailored to the type of positions advertised.

Advertisements should give a broad description of knowledge and competencies required, and should

not be so specialised as to discourage suitable applicants. Employers should include a description of the working conditions and entitlements, including career development prospects. Moreover, the time allowed between the advertisement of the vacancy or the call for applications and the deadline for reply should be realistic.

Selection
Selection committees should bring together diverse expertise and competences and should have an adequate gender balance and, where appropriate and feasible, include members from different sectors (academic and non-academic, including enterprise) and disciplines, including from other countries and with relevant experience to assess the candidate. Whenever possible, a wide range of selection practices should be used, such as external expert assessment and face-to-face interviews. Members of selection panels should be adequately trained.

3.3 Appropriateness of the infrastructure of the participating organisations

The participating Organizations, both academic and non-academic, were chosen based on

- the primary goals of the BigDAPHNE project;
- the proven ability of the research teams to collaborate, shown in past projects;
- the affinity of the non-academic organizations with the entire training program of the project; and
- the strong desire of the non-academic partners to create links with fundamental research, as well as to train the ESRs through internships and provide mentoring for their future careers.

To this end, the five public universities and research institutions (UNIPI, CEA-ParisSaclay, UNISA, 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, which had a major impact in paving the way to the Higgs boson discovery at the LHC in 2012. Recently, the collaborations of UNISA with UNIPI and AUTH on Diboson physics and on advanced techniques on fast, online data analysis and selection (IAPP FP7 EU-project FTK), are making pioneering advances in particle physics research and instrumentation. Furthermore, addressing the most important questions on Cosmology and Gravity (Dark Matter and Gravitational waves) by the strong research teams of this network [(UCL, CEA, UNISA) and (UNIPI, AUTH) respectively], enhances the interdisciplinary training of the network's ESRs and brings at the heart of the basic research the need for advanced techniques in Big Data handling and analysis, which is a common challenge in all three fields.

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 internships addressing the needs for Big Data tools and analysis techniques. To this end, the Software Carpentry Foundation (SWC) and the AUTH IT-Center, have long-standing experience with courses on software and Big Data tools, while SWC are world leadrs in software training workshops for scientists and in forming new software training instructors. OPEN-Lab, INFN, INAF, CNR and companies operating in diverse areas, like Thales, ENEL, NCC, TfL, have strong IT divisions and manage large data analytics projects, while companies like Semblent, as providers of data analytics services to private clients, are well-poised to provide training and internships for the project. The SoBigData project (a consortium led by the University of Pisa and CNR) and OPEN-Lab (a consortium of CERN and world-leading ICT companies, like ORACLE, INTEL, Rackspace) will also provide expert training and internships on data mining, data management and relational and non-relational databases.

ENEL, TfL, NCC and Thales are interested in and have the experience and background to address Big Data related aspects like: applications of Big Data in energy management and optimization, Big Data services, data protection and security, information assurance and protection of cyber attacks, Optimization procedures, predictions and preventions in data handling, to mention a few.

Finally, all the ESRs of the project will have access both to local clusters and to the largest computing

systems available in Europe and the US (like the entire LHC Grid Cumputing 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, UNISA, AUTH, UCL) have also access to national or university-wide computer farms. Thus, the necessary computational resources of the project are fully secured.

3.4 Competences, experiences and complementarity of the participating Institutions

Within the thematic work packages 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 scrutiny of the ATLAS data collected in RUN1, from 2009 to 2012, and at the start of the RUN2, in 2015. 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 a ultra fast online selection of such kind of final states based on programmable electronics and associative memories. On this ground a strong cooperation exist with the AUTH and UCL groups, which were already partners in the EU funded project IAPP (FP7, grant agreement n.324318). 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 imbalanced b-jet system in the final state define a shared research between UCL and UNIPI. UNISA, 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 milestone reached by ATLAS, and the CMS experiment, at LHC leading to the assignment of the Nobel prize for physics 2013. 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 SM. The successful strong commitment to the Higgs program of some of the teams of BigDAPHNE has a demonstration on the grant ARTEMIS of FP7, successfully lead by Dr. R. Nicolaidou from CEA, and involving AUTH, UCL and PISA. 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 out in RUN1 with first raw involvement of the UCL and UNISA 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. The experience of the UNISA team on strong gravitational lensing and the long standing collaboration with Dr. Scaramella of INAF are the ground for the development of shared projects between this team, joining now EUCLID, and the CEA and UCL groups. Finally, the PISA and AUTH groups have also strong links and complementary expertise on the VIRGO project, being Pisa a member of the collaboration, involved with detector operation in addition to data analysis, and Dr. N. Sterg from AUTH a component of the VIRGO Theory Advisory Board.

STOP PAGE COUNT – MAX PAGES (SECTIONS 1-3)