



# BigDAPHNE: BIG DAta in PHyysics NEtwork

Network for an Innovative Doctorate Training in  
Fundamental Physics Research and Big Data Handling

a proposal for an Innovative Training Network (ITN) for a European Joint  
Degree (EJD)

N.Konstantinidis, R.Nikolaidou, C.Petridou, C.Roda, S.Spagnolo

# The opportunity: EU-ITN

EU offers the possibility to fund projects that propose an “Innovative training” for PhD students

According to EU the purpose of this innovative training is:

- Create a new generation of researches better formed to face innovation
- Extend the traditional academic research training increasing transferable competences, inter-disciplinarity, internationality
- Improve the interaction between academic and non-academic sectors

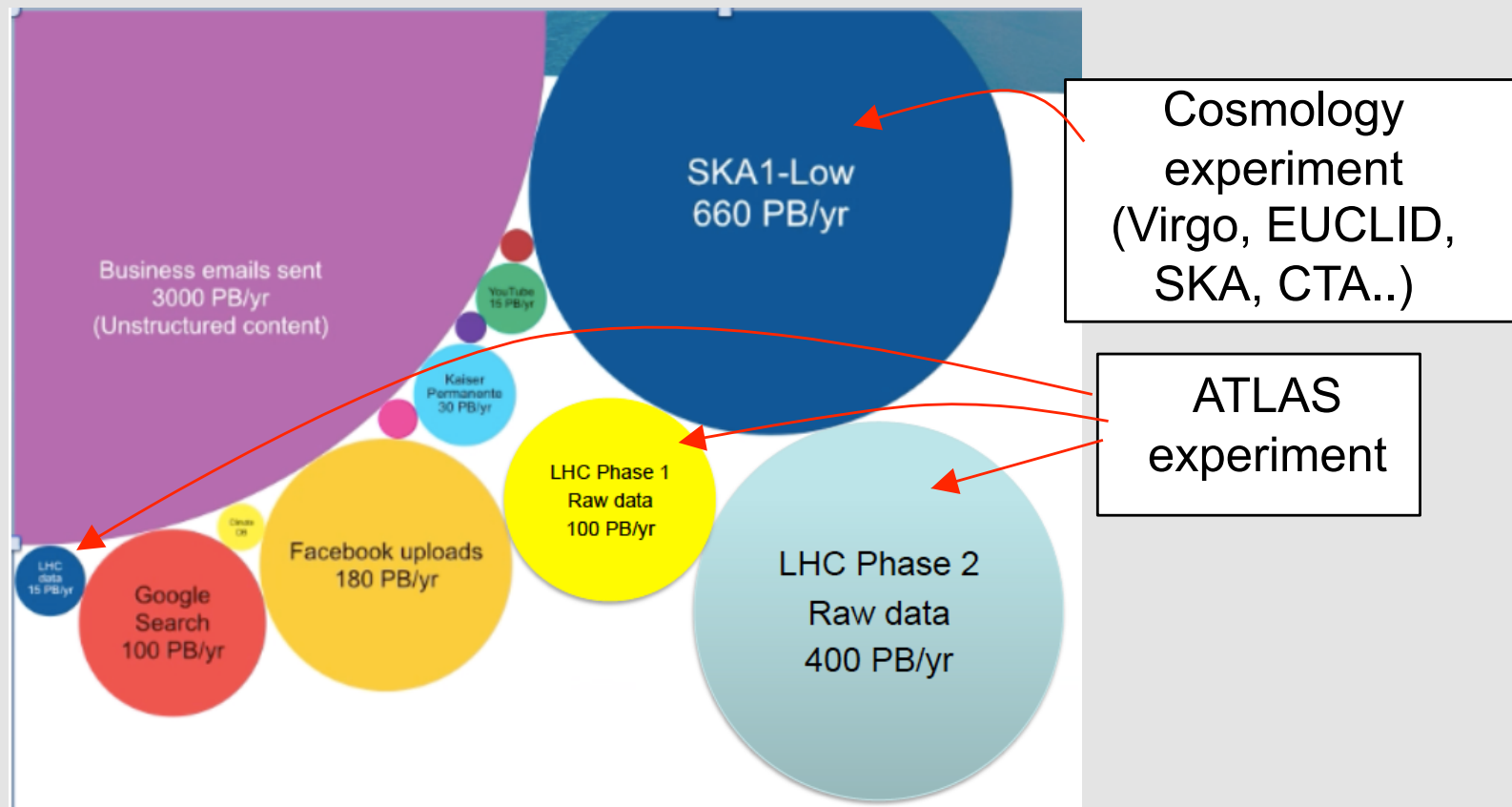
# The idea of our project

# BIG DATA CHALLENGES AND FUNDAMENTAL RESEARCH

(Particle Physics and Cosmology)

- The last years have seen a huge increase in the possibility to accumulate data from different sources (footprints on internet, social networks, sensors, market, ...)
- The possibility to use this information has big potential in many fields from fundamental research up to development of government strategies
- Development of tools to collect, store, analyze this huge size data set has been very prolific both in companies and in the research environments

We think we have a very good environment to extend the training of Physics PhD students in Big Data technologies and challenges



Students working on PhD thesis in Particle Physics and Cosmology are exposed to data sizes very similar to those produced for example in Google searches or FaceBook uploads

**Having a good training in good SW practice, in data science tools in physics research environment and in non-academic environment would produce a very good scientist profile**

# Implementation

# BigDAPHNE

## Research focus



- PhD in Physics
  - Particle Physics - ATLAS-LHC
  - Cosmology - VIRGO, EUCLID, SKA (tbd)
- Double PhD obtained in 2 institutions of the network

## Training focus



- BigData technology and tools
- Internships in companies
- Good SW practices and experience in communicating SW

# Training focus: BigData technology and tools

## Summer schools on Big Data techniques in non-physics environment

- This school will be organized by the Computer Science Department of University of Pisa
- Program will be based on the content of the Master in Big Data analytics and social mining: <http://www.sobigdata.eu/master/bigdata>
- In this case focus is on applications and methods used in non-physics environment

## Summer schools on Big Data techniques in Particle Physics and Cosmology

- CERN and OpenLab has wide experience in organization of summer schools/workshops /courses in computing/sw in particle physics
- Can we build on this to organize a school more oriented to the acquiring/storing /analyzing large data sets ?
- Synergy between tools and techniques used in cosmology and particle physics

## Summer school in statistics (theory, tools, hands-on)

- Thessaloniki in collaboration with INFN
- Several training organized
- Computing center/GRID infrastructure available



# Training focus: seminars/workshops

Each year we will have a project meeting

- Exploit this opportunities to have more contributions on training from companies (oriented to their tool and their problematic) and Computer Science oriented institutions
- This is the moment to gather proposals for this shorter trainings
- We would like to design the training program enforcing cross-environment contributions and multi-disciplinarity

# Training focus: Internships

## **Internship experience has a fundamental role in training**

- Students will spend 3-4 months in one company or in a non-physics institutions working on a Data science project proposed by the institution where the internship is held
- Data science projects can cover tool development or data analytics in non-physics environments
- A member of the institution is required to:
  - supervise the student during internship
  - act as mentor during all PhD on the data science related works

Present contributors to the internship program: Thales (France), ENEL (Italy), Semblant (UK)

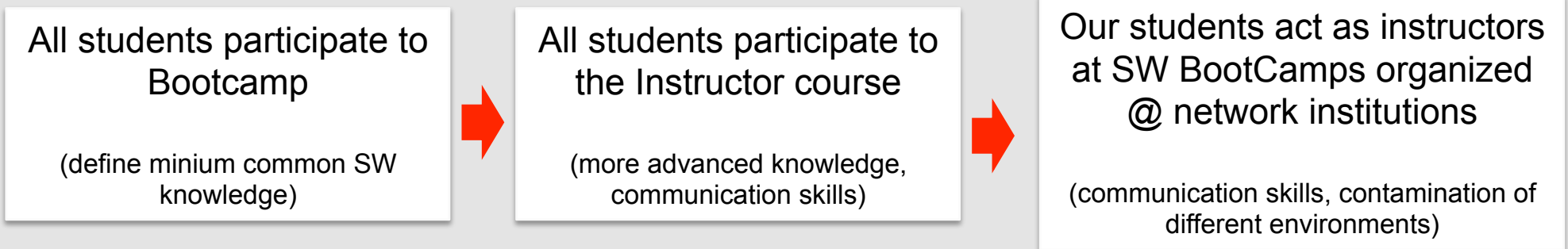
Other possible contributors: CERN/OpenLab (Switzerland), SSI (UK), RedHat (Italy)

# Training focus: SW good practice and communication

A fundamental ingredient in the training program is teaching good practice in SW development

Use SW Carpentry approach: SW Bootcamps (2/3 days hands on courses with instructors from various research environments)

## IN OUR NETWORK – a 3 step training:



Present contributor: SW Carpentry  
More contributors on this point are welcome: SSI (?)

# The project in a nutshell

- 15 PhD students
- PhD thesis on Particle Physics or Cosmology experiments related to use of BigData
  - Particle Physics: ATLAS
  - Cosmology: EUCLID, Virgo, SKA (?)
- Each student will:
  - obtain a double degree from two universities in the network
  - Follow training on BigData retrieval/handling/analytics
  - Follow training on good Software practice
  - do an 3-4 month internship in a company/institute on a BigData project

# Composition

## **Academic for Double Degree**

Thessalonikis University – C. Petridou - Greece – ATLAS, Virgo

University College London – N. Konstantinidis – UK – ATLAS, EUCLID

Universita` di Lecce – S. Spagnolo - Italy – ATLAS, EUCLID

Comm. a l'energie atomique - R. Nikoloaidou – Saclay, France – ATLAS, EUCLID

Universita` di Pisa – C. Roda – Italy – ATLAS, Virgo

## **Partners for training and/or internships**

Thales – B. Resende - France

ENEL – M.Masotti - Pisa, Italy

Semblent – H.Howard – London, UK

Software Carpentry - J. Duckles – Oklahoma

Software Sustainability Insititute - N. Chue Hong – Edinburgh, UK

CERN/OpenLab – A.Di Meglio – Geneva, Switzwerland

INFN – G.Carlino, Napoli Italy

SoBigData CNR, Universita` di Pisa – F.Gianotti, D.Pedreschi – Pisa, Italy

# Logistic

# Approximate Timing

## Indicative timetable for this call

Publication of call	<i>15 October 2015</i>
Deadline for submission of proposals	<i>12 January 2016 at 17:00:00, Brussels local time</i>
Evaluation of proposals	<i>March 2016</i>
Information on the outcome of the evaluation	<i>June 2016</i>
Indicative date for the signing of grant agreements	<i>September 2016</i>

- The project – if approved – last from June 2016 → June 2020
- June 2016 – June 2017
  - Organization of the network, Recruitments of students
- June 2017 – June 2020
  - Schools, thesis, internships...

# Funding and burocratic roles

Funding is based on the number person-month of the project. We will require 540 person-months (15 PhD students)

For each month:

- PhD student salary: 3110 euro / person-month
- research, training and network fund: 1800 euro/ person-month
- management and overheads: 1200 euro/person-month

**Beneficiaries** are organisations that are full partners of a network and are signatories to the Grant Agreement. They contribute directly to the implementation of the research training programme by appointing, supervising, hosting and training researchers. They may also provide secondment opportunities. Beneficiaries take complete responsibility for executing the proposed programme and other requirements of the project.

**Partner Organisations** are not signatories to the Grant Agreement and do not employ the researchers within the project. Partner organisations provide additional training and/or host researchers during secondments.



# Steps for partners 1/2

Each institution that agree to participate as partner need to:

- Retrieve the Participant Identification Code (PIC) :

<http://ec.europa.eu/research/participants/portal/desktop/en/organisations/register.html>

- Provide a letter that describe briefly the type of involvement the company/institution will have in the network

Each partner organisation must **include an up-to-date letter of commitment in Part B.7 of the proposal** to demonstrate their real and active participation in the proposed network. The expert evaluators will disregard the contribution of any partner organisation for which no letter of commitment is submitted. The precise role of each partner organisation should also be clearly described in the proposal. There is no pre-defined number of partner organisations in a project.

# Steps for partners 2/2

For **partner organisations**:

Partner Organisation Legal Name	
General description	
Key Persons and Expertise	
Key Research Facilities, Infrastructure and Equipment	
Previous and Current Involvement in Research and Training Programmes	
Relevant Publications and/or Research / Innovation Product	Max. 3

# How do we go on ?

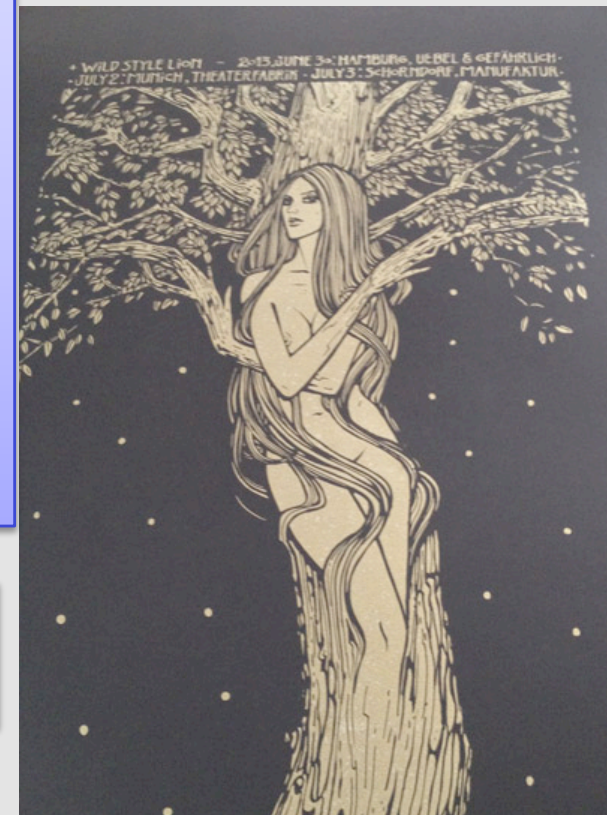
Today we will see a short presentation from each potential partner institution

Need to clarify details of the participation of everybody in a quite strict time





















Please ask questions and make proposal for your contribution

We hope to make BigDaphne to grow old ☺

Need to define the interest to commit and the role of each participant approximately by November 16th



# Contacts

Name	Email	Organization	▲	Chat Name
 Riccardi Juri	juri.riccardi@enel.com	ENEL		juri.riccardi
 Masotti Matteo (E&R RIC)	matteo.masotti@enel.com	ENEL		masotti6913
 t.kitching@ucl.ac.uk	t.kitching@ucl.ac.uk	EUCLID – UCL		tom_d_k
 Francesco De Paolis	Francesco.DePaolis@le.infn.it	EUCLID – UNISA		francesco.depa
 Alberto.Di.Meglio@cern.ch	Alberto.Di.Meglio@cern.ch	OpenLab		shinsenai
 Eric Grancher	Eric.Grancher@cern.ch	OpenLab		
 Roberto Garbelotto	rgarbelo@redhat.com	RedHat		
 Rosy Nikolaidou	Rosy.Nicolaidou@cern.ch	Saclay Laboratoire		rosynikolaidou
 hhoward@semblent.com >> Harry Howard	hhoward@semblent.com	Semblent		hpdhoward
 sforster@semblent.com >> Sam Semblent	sforster@semblent.com	Semblent		
 fosca.giannotti@isti.cnr.it	fosca.giannotti@isti.cnr.it	SoBigData		
 Dino Pedreschi	pedre@di.unipi.it	SoBigData		
 Neil Chue Hong (SSI)	N.ChueHong@software.ac.uk	SSI		
 execdir@software-carpentry.org >> SCF Executive ...	execdir@software-carpentry.org	SW Carpentry		jduckles
 Greg Wilson	gwwilson@software-carpentry.org	SW Carpentry		
 Bernardo Resende	randebor@gmail.com	Thales		ber.resende
 stefania spagnolo	stefania.spagnolo@le.infn.it	Universita` Sannio – Lecce		stefspagnolo
 nikos Konstantinidis	n.konstantinidis@ucl.ac.uk	University College London		nkonstant
 Chariclia Petridou	Chariclia.Petridou@cern.ch	University of Thessaloniki		petridou
 niksterg@auth.gr	niksterg@auth.gr	University of Thessaloniki – Virgo		niksterg1