Contact person: Lorenzo Cazon cazon@lip.pt (www.lip.pt)
DRAFT 3: This document is not yet final. It is open to suggestions from all of the actors involved.

"An event that shows the potential and synergies of high-energy physics and astroparticle physics in the highly demanded field of Data Science in the modern society, and create links with the industry."

Data Science in (Astro)Particle Physics And the bridge to industry

1. Intro/Abstract

Data Scientist is one of the most sought-after jobs of the moment. Social Networks, Large Retail companies, Pharmaceutical, consulting and telecommunication companies, all the Fortune 500 companies are hiring Data Scientists. Everywhere, everyone. Data Science is the art of analyzing data sets to find correlations, causal relations, patterns; build hypothesis, assign significances to them, asses the efficiency of an algorithm of finding a signal probability of false positives, assess the efficiencies of finding a signal. Define control samples, simulate and replicate the reality according to a model. Access, store, retrieve data, moderate or extremely large data-sets (Big Data); create automatic tools that take decisions,..... Data analysis within high energy physics and astroparticle physics is Data Science. The idea is to create an event (a school with a workshop at the end) to show to students, PhDs and open to young post-docs that coming to fundamental physics is an career opportunity with huge synergies with the job market and the needs of modern society. This event would be made of a school with courses on data science currently used in HEP (High Energy Physics) and astroparticle physics and a workshop where different companies present their current trends, needs and daily work related to Data Science.

This event also intends to find interlocutors for establishing communication channels with the industry that allow to explore partnerships and joint projects, to compete for international grants, and also to show the capabilities of LIP and (Astro)Particle Physics in general in the field of Data Science. This event will facilitate bringing together the two sides and create the framework to enlarge mutual knowledge of daily work, allowing to create practical synergies from Data Science in fundamental physics into Data Science in industry.

2. Contents

School *"Data Science in (Astro)Particle Physics"* (Extended list of possible topics,TBD)

*** Statistics & Probability:** frequentist, bayes, significance, false positive, false negative, correlation, causality, paradoxes (regression to the mean), etc
* **Data** T**ools:** C++, python, root, R, java, databases, others(?)
* **Machine learning:** neural networks, decision trees, others,
* **Simulations:** Monte Carlo as a core tool in data analysis/data science
* **Computation**: clusters and parallelization
* **A real example in astroparticle physics** ("learn by doing"), practical example where a global end-to-end mini-project is carried out, divided in the above sub-tasks

Workshop *"Data Science in the industry"*

The goal is to have a diversified sample of cases where Data Science is key in modern society outside fundamental science. (Some of the speakers might be even former scientists who shifted their career from fundamental research at some point). Talks shall show the importance of techniques and knowledge related to Data Science/Analysis that is natural within HEP/Astroparticle Physics research in their actual work, with real examples.

COMPANIES:
At this stage, it is important to establish first contact and show of interest from companies and/or speakers from companies. **All docs and ideas are subject to modification according to the feedback we obtain.**
We maintain a general focus:
* e-companies: internet based
* Non e-companies: pharmaceutics/medicine/biology, consulting, financial, engeneering
* Local companies: large companies, and local start-ups
* We'll try to have some international well known companies and speakers.

3. Exploration of LIP-Industry partnerships and joint projects

The School-Workshop wants to serve as seed for a permanent communication channel of LIP as a partner for joint projects, as for instance:

- Education and training in data science at LIP
- Trainee programs LIP, LIP/Industry
- Data analysis, data mining and knowledge building at LIP
- Simulations at LIP
- Computing
- Design of Data acquisition
- Quality control on physical processes
- Others

The event might hold **round tables for open discussions**, or **bilateral meetings with interlocutors from companies**, or have **specific presentations at the school/workshop showing LIP capabilities**. Once the connections are created, and possible synergies are identified there should be further iterations with a **format to be defined** out of the scope of the school & workshop**.**

4. Organization & Logistics

1. Expected attendance

School: ~20
Workshop: Difficult to asses, but definitely more than >>20

2. Venue

LIP-Lisbon Auditorium (Free) Max capacity 80

3. Budget

Expenses:

Non-local speakers tickets and lodging: 1000€/person from Europe. 2000 k€/person from elsewhere
Rental of facilities: 5 days, Free
Coffee break/food: 5 days, ????????

Revenue:

Fee: To be defined if we set some fees for non local (LIP) PhD or Post-docs
Academic sponsorship: LIP, IDPASC, university,
Other public/non-profit sponsorship: Foundations? Public Institutions?
Private sponsorship: must be investigated within local (and global?) companies:

5. Time Plan

January 2017:

Prepare draft of proposal (this document, or part of it) where major objectives and planning is defined.

February-March 2017:

Contact companies for the workshop, get show of interest
Explore sponsorship possibilities with public institutions and companies.

April 2017:

Converge to a first official proposal

July 2017

Define date of the event: Options
- Week of March 12th 2018
- Week of March 18th 2018
- Week of May 14th 2018
- (April is for advisory committee)

Start preparing the school in detail, with concrete contents

August 2018

Contact speakers
Issue invitations
Logistics, facilities
Advertising for students
Open call for other speakers

2018

Spring: The Event