A European Data Science Institute for Fundamental Physics

Maurizio Pierini

CERN Experimental Physics Department

ABSTRACT: In order to facilitate the deployment of modern data science technologies (e.g., Deep Learning) into theoretical and experimental research in high energy physics, we suggest that the creation of a **European Data Science Institute for Fundamental Physics** is included among the recommendations of the European Strategy group. Such an institute would facilitate the development of cross-collaboration and across-border work on general-interest techniques, as well as yield knowledge transfer from and to other scientific communities (astrophysics, cosmology, computer science, etc.) and tech companies worldwide.

High energy physics (HEP) traditionally makes use of advanced Data Science (DS) techniques, ranging from large-scale computing to advanced statistical analysis and machine learning. To a large extent, these techniques are seen by the community as data analysis tools and their usage has been confined to the latest stages of data processing of typical particle physics experiments.

In recent years, important conceptual and technological progresses have contributed to drastically change the landscape of DS, with deep learning emerging as the ubiquitous solution to a large variety of problems (such as computing vision, natural language processing, autonomous vehicles), possibly contributing to anticipate the advent of Artificial Intelligence. On of Deep Learning's biggest novelties is its effectiveness in processing directly raw data, thanks to a better scalability to large-dimension inputs. For HEP, this implies the possibility of employing DS technology in very early stages of data processing, such as trigger & data acquisition, event reconstruction, simulation, etc., as well as in the possibility of running end-to-end applications going directly from event hits to the measurement of theory parameters in a specific model.

Developments in Deep Learning connect with (and could facilitate) future developments in strategic technological sectors, such as exotic computing architectures (e.g., neuromorphic chips), artificial intelligence, and quantum computing, which could all become very relevant to HEP in a near future. Nowadays, developments in these direction are resulting from a strong collaboration between academia and *new economy* tech companies (e.g., Google, Facebook, etc.). For the HEP community to play a role in this development and benefit from it, it might be beneficial to follow a similar cross-discipline approach.

The HEP community was so far a skeptical spectator of the Deep Learning revolution, with only a few individuals developing Deep Learning applications. Nevertheless, first proofs of principle exhibit very encouraging results. Existing efforts (e.g., the CERN OpenLab framework) to investigate these technologies for HEP do involve an active contribution from tech companies. In this kind of collaborations, physicists bring the knowledge domain and companies typically provide hardware and software solution. The missing ingredient for an optimally fruitful collaboration is the insight of professional computer scientists, data scientists and data engineers, that could give to these efforts that interdisciplinary nature that was a key element in similar successful initiatives (e.g., the NYU Center for Data Science and Berkeley Institute for Data Science).

Data Science could help solving many of the challenges of next-generation particle physics detectors (e.g., speeding up data processing and event simulation for the intensity frontier experiments, e.g., HL-LHC and DUNE). Research in Deep Learning and the new technologies that might emerge from it will have a crucial role in the next decade. It is important for our community to acknowledge this concrete possibility and plan accordingly. In this sense, a recommendation by the European Strategy group to create a dedicated European Institute for DS would have a crucial role in facilitating this process. Such an institute could be created at one of the European laboratories or close to a University with a leading DS research community. The Institute should serve as a research center as well as a venue for short- and long-term conferences, workshops, training schools, and hackathons. Such an institute would provide a home to many fragmented R&D initiatives across Europe and put the European HEP Community in the position of competing with similar efforts already ongoing in other parts of the world.

In this respect, a connection of such a center to CERN might offer strategic advantages. Using CERN visibility, it would be possible to attract private resources from tech companies to contribute to the activities of the institute and establish collaborations on scientific and technological fronts.