

## PRESS RELEASE

### How Big is Big Data in Terms of Big Science?

The HEPTech Academia-Industry Matching Event (AIME) focusing on big data applications in science and industry, hosted by the Wigner Research Center for Physics in Budapest, Hungary, on 30-31 March 2015, explored the possible answers to this question.

The aim of the event was to bring together academia and industry to share ideas and potential applications, and to foster collaborations in the newly emerging field of big data and related areas. It discussed some of the challenges that future HEP projects could face when generating and analyzing scientific data and examined new solutions. The topics addressed, such as Storage in Big Data, Connectivity, Cloud Computing, Internet of Things, Medical Applications, Data Visualization, and Analytics attracted 87 participants from 12 countries. Among them were companies like Microsoft, UStream, HP, SAS and Intel Corporation (UK).



This was the first AIME organized jointly by HEPTech and Wigner RCP. The idea for the topic of the event came from Wigner Data Center, 70% of which is working for CERN. “This AIME is a continuation of what we have learnt and what we have experienced from the academic sector: could we use it, could we transfer it to the Hungarian and regional industry? This is our mission and this event is completely in line with it”, says Dr. Peter Levai, Director General of MTA Wigner Research Centre for Physics. His aim is to consolidate under the umbrella of Wigner Data Centre all the efforts in the field of big data in Hungary – research, connections with companies and education of young people. In this respect, the AIME on big data will provide conditions for

establishing a network of interested parties. As a follow-up, Wigner RCP is planning to organize meetings with the industrial partners that attended the event and to discuss joint projects.

Most of the speakers at the forum pointed out that the scientific developments nowadays could only result from collaborative efforts.



“Big science cannot be done anymore by individuals or single institutions. It relies on international cooperation and there are many reasons for this. Big science requires collection of an enormous amount of data. Then you need the hardware infrastructure and the software platforms to analyze this data. Nobody can produce a perfect platform in isolation. The software has to be developed as part of a large collaboration where everybody can contribute with a little piece to make it efficient. To build a meaningful scientific discovery process today, you need a complete infrastructure to pool together the computing power, the data, the intelligence, and the knowledge to produce valuable information. Without the

technology, without the big data, today we could not do big science”, says Alberto Di Meglio, Head of CERN openlab, a public-private partnership programme run by CERN to establish joint technical and educational activities with leading IT companies (<http://openlab.cern.ch>).

He believes that big data and big science cannot be separated because they have evolved together. As the scale of doing science increased in the past 50 years the amount of data collected also increased. Therefore, the technological requirements to process this data also had to evolve. “At CERN we try to bring research and industry together to understand what the common problems are and how we can solve them together. And then each entity will apply the results in their respective domain. In CERN openlab we work with companies on research rather than on commercial collaboration: we investigate technologies, we understand how to make them better, how we can apply these technologies and big data to science and how the companies can apply them to their products and services”, explains Di Meglio.

Many companies like Microsoft are also very active in establishing collaboration with academia to address more efficiently the needs of the researchers.

“Science at the moment is data driven. Microsoft is a data driven company. We are going to data driven decisions and sell a data driven approach to the market, harmonized with the data driven research”, says Barnabas Doczi from Microsoft, Hungary. Microsoft can give a solution to a researcher for various problems – for instance, how to store and transform the data, how to find the hidden relationships

and how to present the results. That is how the company builds the right environment for the data driven research.

Microsoft has collaboration agreements with universities and is looking forward to cooperate with CERN, as a follow-up of this AIME. “The basics of the Kinect technology and Azure machine learning came from the research. Microsoft further developed them to fit better the needs of the researchers. Now, they can use the Azure machine learning, Power BI and SQL Database to speed up their research results - to spend less time to collect, clean and prepare the data, and more time to analyze it and to work with the results”, explains Doczi.

The spirit of collaboration in the context of big data was also shared by Wolfgang Lengert, from the Earth Observation Ground Segment Department of the European Space Agency (ESA).

“We work with CERN to see how we can leverage the assets provided by the European Member-states and use them in the common interest. We want to identify where we have common methodologies that we could exploit, despite of the thematic differences”, explains Lengert.



He believes that for ESA Earth observation applications big data is very small in terms of big science because the volume itself is not the most important. “Our data is measuring the Earth at a high level precision. For instance, the temperature of the ocean is measured with the precision of 0.1 to 0.3 degrees. However, all the measurements themselves outside any context have no meaning. We have to put them together and see what people can do with them for the Earth. What are the changes in the long run and what is the impact of those changes on health, on climate, on the ecosystem? Therefore, I think that collaboration between the ESA and organizations like CERN is a very good approach and we could contribute

significantly to the big data picture”, says Lengert.

The topics discussed during the event proved that no matter how big big data is, it is essential for the further development of big science.

**Written by:**

**Eleonora Getsova, HEPTECH Communication Officer**