

## Big Data Science Center at SSRF The making of a Superfacility

*Tuesday, 16 July 2019 09:10 (35 minutes)*

The rapid development of synchrotron facilities has massively increased the speed with which experiments can be performed, while new methods and techniques have increased the amount of raw data collected during each experiment. While this has created enormous new opportunities, it has also created tremendous challenges for the national facilities and the users. Traditionally, users collect data during their assigned and limited beamtime and then spend many months analysing them. With the huge increase in data volume, this is no longer possible. As a consequence, only a small fraction of this multidisciplinary and scientifically complex Big Data are fully analysed and, ultimately, used in scientific publications. Therefore, in few years, as the Synchrotron Big Data do definitely overcome any conventional data analysis approach purely based on human resources, Users could then be incapacitated to produce meaningful science from their synchrotron experiments. This problem is even more evident in the case of XFELs, where tens of Petabytes are produced and must be analyzed annually. This is unfortunate because synchrotron beam-time is an expensive resource with respect to money as well as time. Secondly, a lack of appropriate data analysis approach limits the realisation of experiments that generate a large amount of data in a very short period of time, and thirdly, the current lack of automatized data analysis pipelines prevents the fine-tuning of an experimental run during a beamtime, thereby further reducing the efficiency of the beamtime potential usage. This effect, commonly known as the “data deluge”, affects the light sources worldwide in several different ways such as fast data collection and available local storage, curation of the data, including data movement and deposition in a database.

In order to address these crucial Big Data challenges that affect synchrotrons world-wide, Prof. Alessandro Sepe is leading the deployment of a novel Big Data Science Infrastructure at the Shanghai Synchrotron Radiation Facility (SSRF), Zhangjiang Laboratory. Here, Synchrotron Big Data are totally integrated with Artificial Intelligence, High Performance Cloud Supercomputing and Real-time remote robotic experiments, in order to create the first-ever World-Class User-Friendly Superfacility, aimed at accelerating scientific discoveries and technological advancements. Here, also non-experts can obtain scientifically meaningful results in real-time from the multidisciplinary science at Large National Scientific Facilities, like SSRF and Zhangjiang Laboratory. This will effectively extend the use of synchrotron facilities

to the largest plethora of scientific disciplines ever, thus dramatically increasing the scientific outcome of the Users at Large Facilities like SSRF, while aiming at supporting all the key national scientific needs in China. This seminar will focus on the solution that the Big Data Science Center at SSRF, Zhangjiang Lab is developing in order to address this Big Data deluge issue indeed, which poses a serious challenge to the scientific future of all the Synchrotron, Neutron and XFEL large facilities worldwide.

In order to address this challenge, a highly coordinated scientific and technological initiative, bridging state-of-the-art science with the most advanced technology, is required. This seminar will thus cover all the most recent efforts in Big Data Science applied to Synchrotron facilities, as they are implemented at the Big Data Science Center at SSRF, Zhangjiang Laboratory, which represents the very first cornerstone for the creation of a Superfacility at SSRF, Zhangjiang Laboratory.

**Primary author:** Prof. SEPE, Alessandro (Big Data Science Center, Shanghai Synchrotron Radiation Facility (SSRF), Zhangjiang Lab)

**Presenter:** Prof. SEPE, Alessandro (Big Data Science Center, Shanghai Synchrotron Radiation Facility (SSRF), Zhangjiang Lab)

**Session Classification:** 大会报告 I

**Track Classification:** 科学计算技术与平台