

Growth in the data sciences

Another office for the Swiss Data Science Center will be established at the Paul Scherrer Institute PSI. To this end, the ETH Board has approved an increase of five million Swiss francs in the budget of the strategic focus area Data Science. The main aim of this expansion is to help improve the evaluation and processing of the growing amounts of data from large and complex research infrastructures, sensor networks, and databases at PSI and the other three federal research institutes, Empa, WSL, and Eawag. The resources and expertise will be available to all institutes in the ETH Domain.

More precise and modern measurement methods mean more detailed insights into the world that surrounds us. But they also mean that a lot more data is generated. "Processing and evaluating the data is becoming an ever greater challenge," says Gerd Mann, overall head of IT at PSI.

The Swiss Data Science Center (SDSC) supports the ETH Domain with expertise and new methods such as machine learning and artificial intelligence to face the challenges encountered in research problems requiring complex data processing. The SDSC was created in 2017 as part of the strategic focus area Data Science and up to now has been located at ETH Zurich and the École Polytechnique Fédérale de Lausanne EPFL. A third site will now be set up at PSI in Villigen in the coming years. "This new unit will help further bridge the gap between data scientists and domain scientists while addressing the exploding growth of scientific data collected by the large-scale research infrastructures in Switzerland", says Olivier Verscheure, SDSC Director.

Another aim is to expand the existing cooperation between PSI and the Swiss supercomputer center Centro Svizzero di Calcolo Scientifico CSCS.

Data explosion – an opportunity for science

Estimates indicate that over the next four years the amount of data generated annually at PSI alone will increase from the current level of around 3.6 petabytes (= 3.6 quadrillion bytes) to more than 50 petabytes. One reason for this is the planned upgrade of the Swiss Light Source under the project name SLS 2.0. During the same period, the X-ray free-electron laser SwissFEL will be going into regular operation with additional beamlines, and thus new, even more complex detectors will be contributing to the flood of data.

"It is not only PSI that is facing the challenge and the opportunities of the growing amount of data, but also other research areas within and outside the ETH Domain," Gerd Mann stresses. Today, wherever researchers are investigating complex systems, measurements are generating more – and more complex – data. This also applies to the life sciences and environmental sciences, where much of the work involves analysing images and videos. For example, high-resolution video recordings can produce more than seven terabytes of raw data per hour.

Text: Paul Scherrer Institute/Brigitte Osterath

Contacts

Dr. Olivier Verscheure
SDSC Director
olivier.verscheure@datascience.ch

Dr. Gerd Mann
Head of IT
gerd.mann@psi.ch