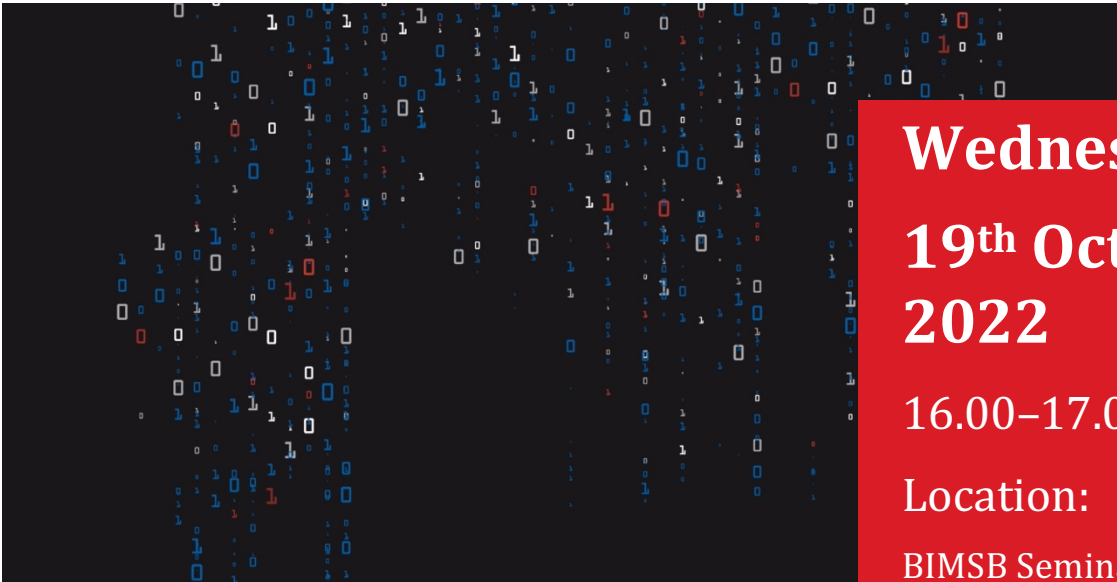


HEIBRIDS LECTURE SERIES

LECTURE SERIES



Wednesday

**19th October
2022**

16.00–17.00

Location:

**BIMSB Seminar Room, 2nd floor,
Hannoversche Str. 28**

Processing of (Scientific) Data in the Age of Data Science

Speaker: Johann-Christoph Freytag, HUB/ECDF

Abstract

My talk addresses different aspects of Data Science to emphasize that Data Science comprises of more than machine learning (ML).

In the first part of the talk, I outline how research in the natural sciences has changed over the last centuries leading to a data driven science approach as we observe it today. I describe a specific example in the astronomy domain exemplifying the new approach to scientific research. Based on this historic view I discuss how to define the term data science more precisely. Using massive amounts of data for research and business motivate scalable data processing platforms such as Hadoop, Spark, or Flink, which only become true due to dramatic changes in computer hardware. In the third part of the talk, I discuss a few important challenges in the area of data cleaning/cleansing to show that this task is one of the most important within data science. In the last part of my talk, I focus on challenges of privacy when managing and manipulating personal data in domains such as in the health/medical domain or business (customer) data. I outline the challenges and briefly describe possible solutions that still allow scientists to use personal data in a useful manner.