FAIR Data Towards New Insights into Materials

Claudia Draxl
Physics Department and CSMB, Humboldt-Universität zu Berlin, Berlin, Germany
Email: claudia.draxl@physik.hu-berlin.de

ID: PLN02
Plenary Talk
Aug. 23, 9.40 - 10.20

Data-centric approaches are becoming an integral part of our research in complementing established traditional paradigms. Given the variety of possible applications in materials science, the potential for gaining new insights through artificial-intelligence (AI) is enormous. The vast amounts of research data produced every day in the field represent a 21^{st}-century gold mine. How can we turn these data into knowledge and value? A FAIR (Findable, Accessible, Interoperable, and Re-usable) data infrastructure plays a decisive role, because this gold mine is of little value if the data are not comprehensively characterized and made available. Only then, data can be readily shared and explored through statistical analysis and machine learning. Making data Findable and AI Ready (another interpretation of the acronym) will change the way how science is done today. With selected examples, I will show how Big Data from either computational or experimental work can be used to find trends and patterns, which would not be possible from single investigations. Thereby, I will also address the challenges when data potentially come from different sources and show how machine learning can be used for error quantification and data augmentation. I will ask how we can make use of existing data collections to find materials with desired properties and will show how ML can be combined with established methodology toward nonlinear modeling and classification.

Keywords: Artificial intelligence, data-centric science, data infrastructure, FAIR data, big data