

Materials Data Curation and Delivery at Scale: Materials Platform for Data Science

Evgeny Blokhin

Materials Platform for Data Science LLC, Tallinn, Estonia

Tilde Materials Informatics, Munich, Germany

The importance of high-quality materials data for humankind is hard to overestimate, yet the scientific community has not been doing enough to systematize the vast swatches of data generated by the research groups all over the world. In my contribution I will elaborate on my role in the PAULING FILE project for the data curation. The PAULING FILE is the private long-running international initiative aiming to cross-check and link the phase diagrams, physical properties, and crystalline structures within the same database via the distinct phase concept, as published in the world peer-reviewed literature since 1891. It backs up such commercial products as MedeA from Materials Design, PCD & APD from ASM International, Springer Materials from Springer, PDF from ICDD, and many others. Recently I have also launched an internet platform called Materials Platform for Data Science, presenting all the PAULING FILE data online. In 2022 it was already visited by thousands of researchers and engineers monthly. I will also present my in-house data-intensive materials research, the typical data-mining scenarios of my industrial customers, and some selected open-source software contributions which I made recently in the field of materials informatics and computer science.