MATERIALS PLATFORM FOR DATA SCIENCE: FROM BIG DATA TOWARDS MATERIALS GENOME

Blokhin E.[a,b] and Villars P.[c]

[a]*Materials Platform for Data Science, Sepapaja 6, 15551, Tallinn, Estonia, e-mail: eb@mpds.io*
[b]*Tilde Materials Informatics, Straßmannstraße 25, 10249, Berlin, Germany*
[c]*Material Phases Data System, Unterschwanden 6, 6354, Vitznau, Switzerland*

Since 1993 the PAULING FILE project team led by P. Villars (Switzerland) excerpts the materials data from the peer-reviewed literature in materials science. As of 2019, nearly 300’000 publications have been processed, and about 450’000 crystal structures, 70’000 phase diagrams, and 900’000 other property sets were extracted. All data have been systematized and critically evaluated. [1] Today PAULING FILE backs up such commercial products as Springer Materials, MedeA Materials Design, and AtomWork Advanced.

The Materials Platform for Data Science (MPDS, www.mpds.io) is an online edition of the PAULING FILE, providing the data in the machine-readable format. Using the PAULING FILE data, we present a holistic vision for all the inorganic compounds. Coupled with the high-throughput ab initio simulations and the machine-learning technique, [2] our vision provides a key to the discovery of materials genome, playing a role of the periodic table for the entire set of materials.

References
1. E. Blokhin, P. Villars, Handbook of Materials Modeling, 2nd ed. by S. Yip, Springer, 2018, In Section "Materials Informatics"
2. E. Blokhin, P. Villars, Quantitative trends in 8 properties of inorganic compounds via machine learning, 2018, submitted