NOMAD Electronic Laboratory Notebook (ELN) to generate Findable and AI-ready Crystal Growth Data

Ta-Shun Chou[1], Andrea Albino[2]*, Hampus Näsström[2], Jose Marquez Prieto[2], Florian Dobener[2], Lauri Himanen[2], David Sikter[2], Mohammad Nakhaee[2], Amir Golparvar[2], Andreas Popp[1], Sebastian Brückner[1,2], Martin Albrecht[1], Markus Scheidgen[2], Claudia Draxl[2,3].
*lead presenter: andrea.albino@physik.hu-berlin.de
1 Leibniz-Institut für Kristallzüchtung (IKZ), Berlin, Germany
2 Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany
3 The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Berlin, Germany

Efficient synthesis protocols for materials with well-defined properties in a reproducible fashion is of paramount importance to materials science. This need is not always fulfilled because it requires control of a large number of experimental details, and it is typically not possible to grasp the full entirety of the relevant parameters with the manual analysis of data. The concept of data-centric science and the development of AI tools promise to model synthesis more reliably and to identify the relevant set of relevant descriptors and their mutual inter-dependencies, or at least their correlations.
The FAIRmat [1] project (fairmat-nfdi.eu) combines all the efforts in designing research data management (RDM) tools for the new data driven approach to the material science [2]. The tool of choice to capture synthesis experiment data is the Electronic Laboratory Notebook (ELN), which is a digital tool used to replace traditional paper laboratory notebooks. NOMAD (nomad-lab.eu) offers ELN functionalities that aim at offering a secure environment to protect the integrity of both data and metadata. The data properties that are available differ from domain to domain. There will be subsets of common properties for each sub-domain, and these subsets form a hierarchy. A bottom-up approach, going from a particular set of experiments to a general description of the similarities recurring in each of them, lead us to envision a common data structure as a standard. For example, experiments and synthesis share a common notion of material, measurement or sample. Furthermore, linking data from synthesis to data from experimental materials science and theory using common metadata schemas and ontologies [3] will create a new level of the science of materials synthesis.
In this talk, we will show how the NOMAD platform can be applied on the whole RDM life cycle in crystal growth and epitaxy meeting FAIR data standards. On the example of data from Gallium Oxide epitaxy [4] we will demonstrate the state-of-the-art ELN features for a synthesis process in NOMAD, focusing both on the underlying data model concepts and on the implementation of a use case,. We will show how structured data facilitates and automatizes data analysis and visualization, and allows to routinely access machine learning tools for process optimization in synthesis [4].
We acknowledge Tamás Haraszti, Holger von Wenckstern, Natascha Dropka, Fabian Zemke, Jonathan Noky, Oliver Bierwagen, Theodore Chang, Markus Kühbach, Sherjeel Shabih, Sandor Brockhauser, Nathan Daelman, Joseph Rudzinski, Luca Ghiringhelli for fruitful discussions.

References

[1] Scheffler M et al. FAIR data enabling new horizons for materials research. Nature. 2022;604:635-642.
[2] Wilkinson M et al. The FAIR Guiding Principles for scientific data management. Sci Data 2016;3,160018.
[3] Gruber T R, A translation approach to portable ontology specifications. Knowledge Acquisition. 1993;5(2):199-220.
[4] Chou, T.-S.et al. Toward Precise n-Type Doping Control in MOVPE-Grown β-Ga2O3 Thin Films by Deep-Learning Approach. Crystals. 2022;12(1):8