A distributed platform for AI-ready solar cells research data

José A. Márquez[1], Hampus Näsström[1], Michael Goette[2], Fabian Schöppach[1], Yaru Wang[1], Sherjeel Shabih[1], Lauri Himanen[1], Markus Scheidgen[1], Thomas Unold[2], Eva Unger[1,2], Christoph T. Koch[1], Kevin Jablonka[3,4] and Claudia Draxl[1]

[1] Department of Physics and CSMB, Humboldt-Universität zu Berlin, Berlin, Germany
[2] Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany
[3] Center for Energy and Environmental Chemistry Jena, Friedrich Schiller University Jena, Jena, Germany
[4] Helmholtz Institute for Polymers in Energy Applications Jena (HIPOLE Jena), Lessingstrasse 12-14, 07743 Jena, Germany
jose.marquez@physik.hu-berlin.de

The development of advanced materials for solar cells remains a slow and resource-intensive process, hindered by the vast complexity of possible chemical compositions and device architectures. To address this challenge, leveraging data science and FAIR data principles is essential for accelerating discoveries and optimizing designs.

This contribution introduces the open-source NOMAD ecosystem (https://nomad-lab.eu),[1] developed by the FAIRmat consortium (https://fairmat-nfdi.eu), as a solution to these challenges. NOMAD enables the collection, organization, and analysis of materials science data, making it a central hub for advancing solar cell research. Central to this effort is the integration of curated datasets, such as the perovskite solar cell database,[2] which provides a rich source of structured and interoperable data for understanding material trends and facilitating data-driven innovation.

Beyond centralized data repositories, the NOMAD ecosystem supports a federated infrastructure concept, enabling distributed instances, each known as NOMAD Oasis. These local installations provide laboratories with customizable tools, including adaptable electronic lab notebooks (ELNs), to seamlessly capture, process, automate, and transfer AI-ready data and metadata. This infrastructure bridges individual lab-scale digitalization efforts with global collaboration, fostering a unified and FAIR ecosystem for materials research data.

To illustrate these capabilities, I will present several examples of how NOMAD services are being utilized to advance solar cell research. First, by leveraging publicly available data, we use network analysis of charge transport layers in perovskite solar cells (Figure 1) to trace the evolution of materials selection throughout the history of this technology. Next, I will showcase how we are enriching the solar cell research ecosystem through the digitalization of individual laboratories, creating a growing interconnected network that spans multiple disciplines within materials science. Finally, I will highlight our ongoing efforts to automate data curation from scientific literature using language models, a step that significantly accelerates the integration of high-quality data into the NOMAD ecosystem.

Figures

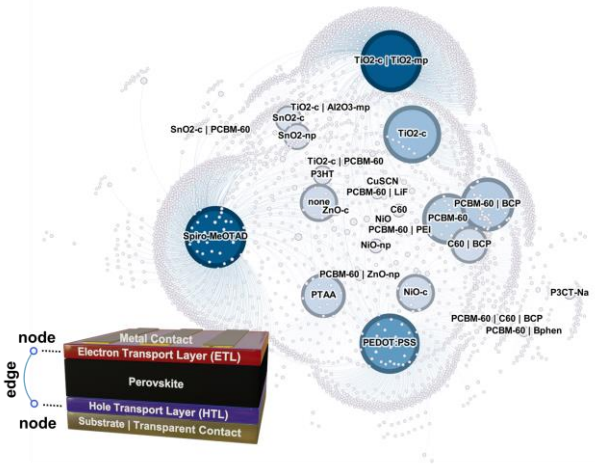

Figure 1. Transport layers materials network in perovskite solar cells. Every node in the network represents a stack of transport layers, either hole or electron. A sketch of a simple perovskite solar cell stack is provided as an explanatory inset, illustrating the nodes and edges.

References

[1] M. Scheidgen *et al.*, 'NOMAD: A distributed web-based platform for managing materials science research data', *Journal of Open Source Software*, vol. 8, no. 90, p. 5388, 2023, doi: 10.21105/joss.05388.

[2] T. J. Jacobsson *et al.*, 'An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles', *Nat Energy*, vol. 7, no. 1, Art. no. 1, Jan. 2022, doi: 10.1038/s41560-021-00941-3.