AI and data-driven infrastructures for workflow automation and integration in advanced research and industrial applications

Tommaso Forni[1,2], Mario Vozza[1,2], Fabio Le Piane[1,3], Andrea Lorenzoni[1], Matteo Baldoni[1] and Francesco Mercuri[1,*]

[1]DAIMON Lab, CNR-ISMN, via P. Gobetti 101, Bologna, 40129, Italy
[2]Department of Control and Computer Engineering, Polytechnic University of Turin, Corso Castelfidardo 34/d, Turin, 10138, Italy
[3]Department of Computer Science and Engineering, University of Bologna, Viale del Risorgimento 2, Bologna, 40136, Italy

Abstract

The use of AI and data-driven technologies and infrastructures for innovation and development of advanced research and industrial applications requires a strong degree of integration across a broad range of tools, disciplines and competences. In spite of a huge disruptive potential, the role of AI for research and development in the context of industrial applications is often hampered by the lack of consolidated and shared practices for transforming domain-specific processes for generating knowledge into added value. These issues are particularly striking for small-medium enterprises (SMEs), which must adopt clear and effective policies for implementing successful technology transfer paths for innovation. The activities of the DAIMON Lab of the CNR-ISMN focus on the design, development, implementation and application of integrated modelling, data-driven and AI methods and infrastructures for innovation in hi-tech applications. Our approach is based on the development of horizontal platforms, which can be applied to a broad range of vertical use-cases. Namely, we target the realisation of high-throughput workflows, related to specific domains and use cases, which are able to collect and process simulations and/or physical data and information. The implementation of an interoperable integration framework is a prerequisite for further application of AI tools for predictivity and automation. With a strong focus on the development of key enabling technologies (KETs), such as advanced materials, the approach pursued is extended to a broad range of application fields and scenarios of interest in industry, including electronic and ICT, advanced and sustainable manufacturing, energy, mobility.

Keywords

Workflow automation, Semantic technologies, Data-driven integration, High-performance computing

1. Introduction

The industrial and academic R&D landscape still largely relies on a trial-and-error approach to innovation and improvement, which can be time-consuming, expensive, and ineffective. Machine learning and AI have the potential to revolutionize problem-solving and decision-making in research and, in particular, for innovation in Industry 4.0. Accordingly, recent years have witnessed the increasing role of machine learning and AI in shaping the future of innovation in high added-value applications, in manufacturing and in industry. The lack of systematic improvement in several research and application fields is a major challenge that can be addressed by the integration of knowledge and data and the automation of the innovation process. The impact of data-driven integration technologies can potentially affect multiple value chains and fields of interest in industry applications and related domains, including but not limited to advanced materials and manufacturing, electronics, mobility, environment, and more.

The realisation of data-driven workflows is critical for generating integrated knowledge that can support AI and data-driven methods for prediction and automation. By implementing data-driven workflows, R&D processes and activities can be optimised, identifying areas for improvement and innovation, and promoting collaboration and knowledge sharing across different departments and areas of expertise. However, the implementation of efficient data-integration frameworks in several application domains is still hampered by a manifold of theoretical and technical issues. Indeed, the uptake of digital technologies for innovation requires to face the challenges related to the integration of data-driven techniques with consolidated processes in specific application and industrial domains. The main difficulties are related to the need for broad multidisciplinary expertise and cross-discipline links. In addition, several fields with a huge potential for

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
tommaso.forni@ismn.cnr.it (T. Forni); mario.vozza@ismn.cnr.it (M. Vozza); fabio.lepiane@ismn.cnr.it (F. Le Piane); andrea.lorenzoni@cnr.it (A. Lorenzoni); matteo.baldoni@ismn.cnr.it (M. Baldoni); francesco.mercuri@cnr.it (F. Mercuri)
0009-0003-2138-5069 (T. Forni); 0000-0001-7663-0306 (M. Vozza); 0000-XXXX-XXXX-XXXX (F. Le Piane); 0000-0002-4860-153X (A. Lorenzoni); 0000-0003-2958-1091 (M. Baldoni); 0000-0002-3369-4438 (F. Mercuri)

innovation still lack shared technical procedures, knowledge and standardisation in the adoption of digital and data-driven technologies. The role of fully-digital approaches, from modelling and simulation to AI, must therefore be consolidated for an efficient link to specific value chains. Several recent research efforts tried to address the challenges of data-driven integration for domain-specific applications. For example, Barbella *et al.* [1] introduced a semi-automatic approach to the integration of data from various sources. The proposed methodology relies on a syntactic/semantic merge of data, thus pointing to the role of semantic technologies for integration. Other recent works discuss similar efforts to data-integration and to the realisation of data-driven automated workflows for example in the context of ecological monitoring [2] and microscopy images [3]. A particularly relevant issue concerns the implementation of data-centric frameworks based on computational and simulation data. Despite the huge potential, related to the possibility to generate meaningful data with high-throughput, this approach still faces several challenges, as discussed recently for the automation of workflows in the modelling of sustainable water [4]. The activities of the DAIMON Lab of the CNR-ISMN focus on the design, development and application of high-performance and high-throughput software and hardware frameworks and infrastructures for the physical and data-driven multi-scale modelling of complex systems for advanced technologies (see Fig. 1). The approach of our lab is particularly relevant in the context of Industry 4.0, linking physical models to data-driven technologies for prediction and automation platforms and frameworks. In this paper, we will discuss our approach to the development of data-driven and AI infrastructures for the automation of data-centric workflows, also showing current applications on a broad range of technology sectors of interest for industry.

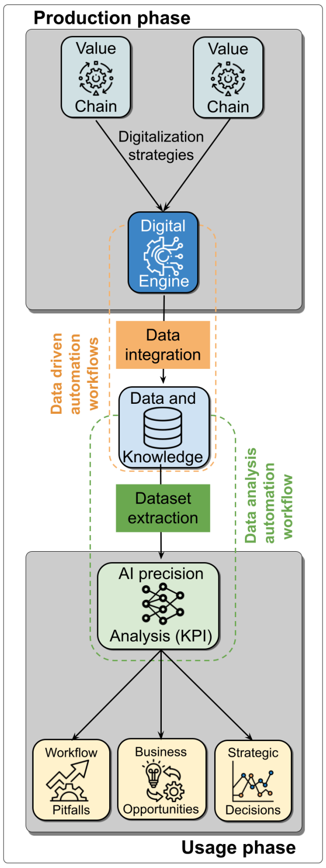

Figure 1: From value chains to added value through data-driven automation strategies

2. Data-driven and automation workflows for research and innovation

One of the major challenges for the adoption of digital technologies for research and innovation is the realisation and implementation of automated data-centric workflows. This challenge is further complicated by the number of potential sources and the increasing volume of available data, generated by research activities or by specific R&D high-throughput analysis tools. Automating the process of data generation and providing native support for data integration and elaboration has therefore become a crucial focus area for both academic and industrial research. In particular, physical modelling and simulation tools have proved their potential in a large number of application fields [5, 6, 7, 8, 9], as they provide a means of exploring complex systems and predicting their behavior. However, these activities often lack clear automation and data-driven integration strategies. Therefore, there is a need to develop automated workflows that can handle large volumes of data and integrate different modelling and simulation tools seamlessly. Similarly, experimental workflows also require automation to improve efficiency and support integration. Furthermore, the lack of frameworks for data interoperability can hinder the integration of data from different sources. This issue is particularly relevant in industrial applications, where data generated from different sources must be aggregated to obtain valuable knowledge and insight for innovation of processes and products. To address these issues, a comprehensive approach to data management would be needed, involving the development of tools and standards for data exchange and interoperability in specific domain applications.

One of the most critical steps for integration is therefore related to the processes involved in the acquisition of data. An efficient integration strategy, interfacing with sources generating data, is essential for the development of a data-driven platform for decision making. Moreover, integration is required also at the knowledge level, which is also defined by the application domains, thus addressing the issue of interoperability at different levels.

In analogy with previous efforts [1, 2, 3], our approach to automation and to the realisation and application of frameworks and infrastructures for AI relies on the design and implementation of data-driven workflows for automating the data integration process. Rather than developing general-purpose platforms, our approach is based on the definition of domain-specific workflows, which are integrated to build integrated data structures and knowledge. Clearly, this approach requires a strong integration between generic data-driven technologies and domain-specific methods and approaches. This integration approach enables the collection and the analysis of data and knowledge at different levels of abstraction. Namely, AI other data-driven tools (expert systems, decision support systems, etc.) can operate on data integration frameworks as services, providing answers to technological queries in terms of predictions, analysis and process automation (see Fig. 2).

One of the key ingredients for the multi-level integration of data and knowledge relies on the adoption of efficient methodological frameworks and technologies for knowledge representation and syntactic and semantic interoperability [10], which can be considered as a key enabler for integration. In particular, ontologies can support and enable semantic interoperability between different systems and applications within or across application domains, by providing a shared understanding of concepts and relationships and supporting the formal encoding of semantic meanings [11]. The formalisation of the intrinsic structure and patterns of a domain of knowledge as a mean for creating encoding of properties and the relationships between different concepts constitutes an invaluable tool for gathering new knowledge from historical data, improving current processes and data collection procedures [12]. For these reasons, semantic platforms can therefore be a key element of unification in the research and development space, when backed by the cooperation among teams of researchers, and make existing research lines more interoperable and shareable [13].

Another key enabling technology for the implementation of integrated data-driven frameworks is the application of high-performance computing (HPC). The implementation of integration strategies on high-performance infrastructures can support innovation by providing the required power to sustain large volumes of data in high-throughput acquisition processes and analysis. In particular, an increase of the overall throughput of integration frameworks can be associated to the improvements in GPU computing technologies and to the development of large-scale data centers. As we will discuss in the next section, the integration approach based on multiple interconnected abstraction levels and corresponding software stacks, and the strong connection with the knowledge and competences of specific application domains are key to develop solutions addressing research and innovation challenges.

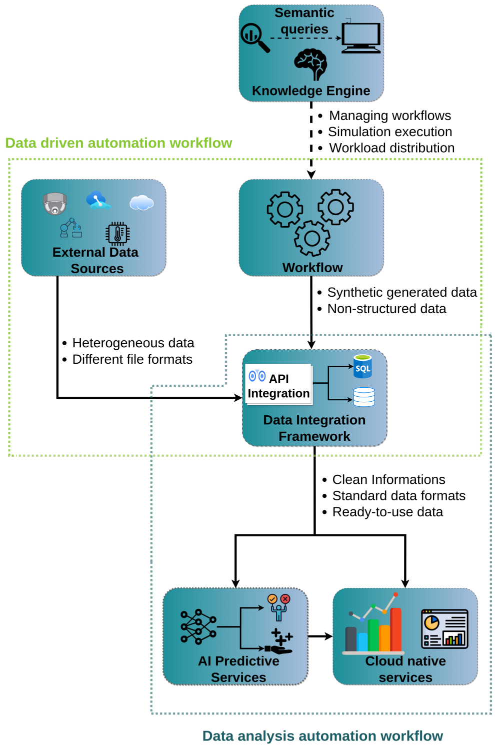

Figure 2: Data Integration Framework for Predictive Analytics and Process Automation. The diagram illustrates the seamless integration of semantic queries, external data sources, and advanced workflows to generate AI-powered predictive services and cloud-native solutions, enabling efficient and effective analysis and automation of complex technological queries.

3. Applications to AI for industry: advanced materials, manufacturing, electronics, energy, smart mobility

The general approach to the development of data-driven and automated workflows and frameworks has recently been applied by our research team to a manifold of different activities. The common trait of this paradigm consists in the integration between a horizontal layer of tools, methods and technologies to vertical use-cases. As stated above, the development of several application fields depends crucially on innovation in a manifold of critical technologies, involving phenomena at different levels and scales, from the very basic constituent of the physical world to processes and products impacting on specific socio-economic sectors. Our work aims at integrating these different levels through digital and data-driven strategies, interconnecting competences and knowledge (see Fig. 3). One of the most strategic KETs for industry and manufacturing is that of advanced materials. In particular, the development of new nanostructured functional materials can enable a wide range of applications in fields including electronics and optoelectronics, energy, health [14, 15, 16]. To address complex structure/property relationships in materials through data-centric approaches, our research group is active in R&D projects on the application of multiscale materials and process modelling. Our approach consists in the development of computational automated data-driven and high-throughput workflows for gaining a better understanding on the materials properties in the context of technology applications. Physical models provide a valuable predictive platform for the design of new materials and processes, correlating the results with available experimental data across a broad range of scales. Recent work demonstrated the potential of this approach in the development of multi-scale functional novel materials for applications for example in optoelectronics (displays, lighting) and for new-generation solar cells [15, 17, 18, 19]. Graphene-based materials, such as nanographenes and graphene oxide (GO) [20, 21, 22], constitute another particularly relevant class of nanostructured low-dimensional materials with a huge potential for the development of new applications in technology. One of the challenges for the uptake of graphene research in industrial applications is, however, related to the translation of lab-scale innovation into technological solutions. The development of high-throughput approaches for automating research on graphene-based materials is therefore a key step for innovation in this field. Basing on computational tools for building meaningful structure/property data about GO samples [23], data-driven approaches can be applied to generate integrated datasets, linking suitable representations of the structure of GO materials to target chemico-physical properties. The ***GrapheNet*** project, carried out by our lab, aims at exploiting AI technologies in the field of graphene research. The main idea behind GrapheNet consists in applying AI and computer vision frameworks commonly used in the analysis of images to graphene-based materials. This approach stems from the quasi-2D morphology of graphene and related systems, which are thus suitable for an image-like representation of structural features and patterns, as done recently in similar work [24]. In particular, we used computational data [23, 25] to train a deep learning model by encoding structural information on graphene samples into a standard image format. The application of convolutional neural networks (CNNs) to the encoded information exhibits remarkable accuracy in predicting the physical properties of graphene samples (average error below 4%) with a gain of several orders of magnitude in computational time with respect to calculations based on physical models. Remarkably, the use of image encodings also outperforms standard machine learning methods in materials science for the representation of structural features. The realisation of an efficient data-driven workflow for the prediction of critical materials properties, for example for electronics applications as in the case of graphene materials, can enable the efficient design of advanced materials and boost the development of applications and products. These results also highlight the potential of a cross-disciplinary combination of competences, as computer vision, image analysis and object detection and computational materials science, which are commonly focused on different application fields.

The integration of data and knowledge must be supported by robust approaches to standardisation and interoperablity. To this end, our recent research efforts were targeted to the application of semantic technologies, with a focus on advanced materials, as a KET for research and industrial innovation. Indeed, the field of advanced materials applications still requires significant efforts for standardisation, integration and interoperability. In this respect, we recently carried out the development of ***MAMBO*** - Materials and Molecules Basic Ontology [26], as a fundamental step for the application of semantic technologies in the field of advanced materials. In addition to providing a domain ontology for materials applications, the development of MAMBO helped us to assess the general requirements for the integration of knowledge in fields where data and information are scattered and standardisation of workflows is still lacking.

The integration of research tools and methods, from high-throughput simulation methods, HPC, semantic and software technologies, aims at providing a paradigm for implementing multi-scale modelling frameworks for advanced materials. This approach is currently being pursued in the framework of collaborative national and international R&D projects (for example, the BIO-SUSHY project, for developing sustainable surface protection by glass-like hybrid and biomaterials coatings [27]). The integration approach allows to connect high-throughput simulations to data-driven technologies, enables multi-scale links for the description of materials properties and provides a basis for predictive and generative platforms for the design and development of functional materials

and applications.
The multi-scale approach finds application in the development of functional devices where the properties of new advanced materials can be exploited. In this context, the DAIMON Lab operates in tight connection with experimental research groups by providing digital platforms for the design and predictive modelling of advanced devices. The development of devices for electronics and optolelectronics can enable a broad range of applications in several advanced fields, such as bioelectronics, renewable energy sources and next-generation solar cells. Our efforts aim at developing data-driven automated workflows for the generation of computational data on physical models of full-scale devices, also incorporating parameters an models at a lower scale. In this context, recent activities include the development of automated workflows for the simulation of electronic devices based on functional molecular materials [28] and the analysis of data and knowledge related to next-generation perovskite-based solar cells [29].
A relevant application field related to industrial innovation is that of mobility. Our integrated approach to tackle innovation challenges in the context of smart mobility applications involves a multi-scale and multi-level perspective. Namely, we address the challenges of smart mobility in current and future scenarios by considering the interlinked data and information from the design of vehicle components to traffic in complex environments. Accordingly, this approach also requires a multisciplinary integration of data and technologies, in strong analogy with other use cases. Recently, we developed a proof of concept called ***SUMOhtms***, an automation and standardisation framework for the generation of scenarios for the simulation of urban traffic based on the SUMO simulation package [30]. By automating workflows, we improved the overall throughput of urban traffic simulations, thus broadening their scope and their potential integration with data-driven and AI frameworks. Automated workflows simulate mobility scenarios by using the open-source geospatial data from OpenStreetMap (OSM) as an input. The entire automated process is implemented and managed by using Docker containers, enabling standardisation and cross-platform interoperability, leading to an overall increase of efficiency of the simulations. Additionally, we also took part in activities involving industrial partners in the field of traffic microsimulations, such as the MoMoTec (Modern Mobility Technological Ecosystem) project, where microsimulations were incorporated into the procedure for solving the Capacitated Vehicle Routing Problem. Current research in this context is focused on the realisation of a digital twin of a self-driving car and its integration into an urban traffic context: the aim is to train Reinforcement Learning algorithms for the active safety of the vehicle, and integrate the trained agent in the context of smart cities, making it capable of communicating with the infrastructure and its environment. In the field of industrial automation frameworks, current projects and activities aim at creating a link between physical models of complex manufacturing components and processes and data-driven technologies. For example, low-code decision support design platforms (DSDP) can enable the design, develop, deploy, and use of AI systems on top of physical models of manufacturing systems. A DSDP therefore implements multi-level digital twins of goods, equipment, parts, and processes, connected to AI services to improve their production and use. Work is in progress to develop application-specific digital twins based on data-integration platforms. Similarly to the other application fields discussed above, the design and implementation of digital frameworks for industry automation requires a strong degree of integration between data sources, physical and data-driven modelling workflows and data and knowledge representation. In addition to the support in the development of industrial processes, this approach can enable automation in the R&D process, thus accelerating industrial innovation.
Most of the work described above is carried out within collaborative national and international networks and initiatives. One of the most crucial aspects of the integration efforts described above consists in the activation of successful paths for collaborative research and training. In this context, the DAIMON Lab is active in several multidisciplinary initiatives, aimed at implementing this integration approach. Recent initiatives include participation in the national PhD programme in AI (phd-ai.it), participation in H2020 and Horizon Europe projects and networks, targeting a broad range of clusters and participation in Next Generation EU - PNRR activities and projects. A particularly relevant aspect concerns collaboration with industrial partners, in the framework of national and international projects, public-private partnerships and other initiatives. The realisation of the R&D objectives described above in a tight collaboration with industrial research teams and end users has proven a key enabler for the definition of successful technology transfer paths.

4. Conclusions

The automation of digital tools for R&D and the integration with AI and data-driven technologies and infrastructures is crucial for driving innovation and development of advanced applications in research and industry. However, the lack of consolidated and shared practices, particularly in small-medium enterprises (SMEs), hampers the potential of AI for research and development. The approach of the DAIMON Lab tries to face some of these innovation challenges by developing horizontal platforms that can be applied to a broad range of vertical use-cases through the implementation of interoperable integration frameworks. Based on multidisciplinary efforts, the integration step is a prerequisite for the efficient application of AI tools for predictivity and automation. This approach enables the potential translation of processes of specific value chains into data-driven workflows for the digitalisation and modelling of complex systems and processes. The use of data-driven and automation workflows is critical for generating integrated knowledge that can support AI and data-driven methods for prediction and automation, and for promoting collaboration and knowledge sharing across different areas of expertise. In turn, this multi-level technological integration can enable the application of AI tools for a broad range of operational processes, including the realisation of predictive platforms, design of processes and products, automation, optimisation, analysis of what-if scenarios, etc.

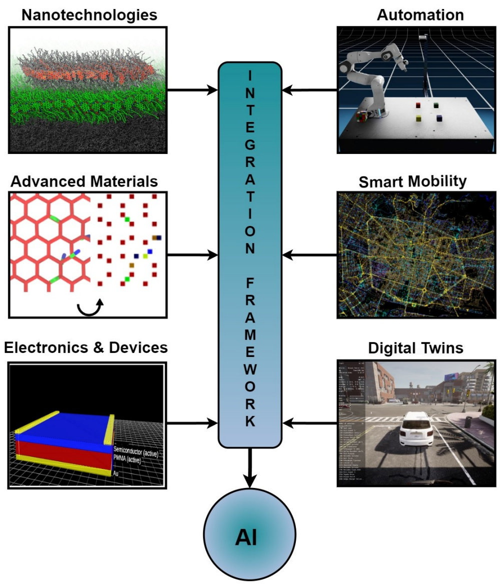

Figure 3: Generic workflow of some of the applications dealt with in the DAIMON Lab. The diagram illustrates how the data produced by the different application fields all converge in a data integration and standardisation framework that prepares them for use by artificial intelligence services

Acknowledgments

The DAIMON Lab acknowledges support from the CNR initiative for the realisation of the AI@ISMN infrastructure and the Italian National PhD program in Artificial Intelligence. We also thank the BIO-SUSHY project of the Horizon Europe programme, the MiTE project for H2 technologies, the MoMoTec project and the PNRR Ecosister project for support. Professor Mauro Gaspari (University of Bologna, Italy) and the aHead research team of Spindox SpA are also gratefully acknowledged for frutiful discussion and ongoing collaboration.

References

[1] M. Barbella, G. Tortora, A semi-automatic data integration process of heterogeneous databases, Pattern Recognition Letters 166 (2023) 134–142. URL: https://www.sciencedirect.com/science/article/pii/S0167865523000132. doi:https://doi.org/10.1016/j.patrec.2023.01.007.

[2] J. Wicquart, M. Gudka, D. Obura, M. Logan, F. Staub, D. Souter, S. Planes, A workflow to integrate ecological monitoring data from different sources, Ecological Informatics 68 (2022) 101543. URL: https://www.sciencedirect.com/science/article/pii/S1574954121003344. doi:https://doi.org/10.1016/j.ecoinf.2021.101543.

[3] A. M. Race, D. Sutton, G. Hamm, G. Maglennon, J. P. Morton, N. Strittmatter, A. Campbell, O. J. Sansom, Y. Wang, S. T. Barry, Z. Takáts, R. J. A. Goodwin, J. Bunch, Deep learning-based annotation transfer between molecular imaging modalities: An automated workflow for multimodal data integration, Analytical Chemistry 93 (2021) 3061–3071. URL: https://doi.org/10.1021/acs.analchem.0c02726. doi:10.1021/acs.analchem.0c02726, pMID: 33534548.

[4] R. A. Vargas-Acosta, L. G. Chavira, N. Villanueva-Rosales, D. D. Pennington, Automating multivariable workflow composition for model-to-model integration, in: 2022 IEEE 18th International Conference on e-Science (e-Science), 2022, pp. 159–170. doi:10.1109/eScience55777.2022.00030.

[5] D. L. Barreiro, J. Yeo, A. Tarakanova, F. J. Martin-Martinez, M. J. Buehler, Multiscale modeling of silk and silk-based biomaterials—a review, Macromolecular Bioscience 19 (2019). doi:10.1002/mabi.201800253.

[6] O. Vitrac, P.-M. Nguyen, M. Hayert, In silico prediction of food properties: A multiscale perspective, Frontiers in Chemical Engineering 3 (2022). doi:10.3389/fceng.2021.786879.

[7] N. Kovachki, B. Liu, X. Sun, H. Zhou, K. Bhattacharya, M. Ortiz, A. Stuart, Multiscale modeling of materials: Computing, data science, uncertainty and goal-oriented optimization, Mechanics of Materials 165 (2022). doi:10.1016/j.mechmat.2021.104156.

[8] A. Lorenzoni, M. Muccini, F. Mercuri, A computational predictive approach for controlling the mor-

phology of functional molecular aggregates on substrates, Advanced Theory and Simulations 2 (2019). doi:10.1002/ADTS.201900156.

[9] J. Fish, G. J. Wagner, S. Keten, Mesoscopic and multiscale modelling in materials, Nature Materials 20 (2021) 774–786. doi:10.1038/s41563-020-00913-0.

[10] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American 284 (2001) 34–43. URL: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.

[11] T. R. Gruber, A translation approach to portable ontology specifications, Knowledge Acquisition 5 (1993) 199–220. doi:10.1006/KNAC.1993.1008.

[12] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B. M. Torres, J. F. Unger, C. Völker, J. Olbricht, A perspective on digital knowledge representation in materials science and engineering, Advanced Engineering Materials 24 (2022). doi:10.1002/ADEM.202101176.

[13] M. T. Horsch, S. Chiacchiera, Y. Bami, G. J. Schmitz, G. Mogni, G. Goldbeck, E. Ghedini, Reliable and interoperable computational molecular engineering: 2. semantic interoperability based on the european materials and modelling ontology, arXiv (2020) 1–25. URL: https://arxiv.org/abs/2001.04175.

[14] D. Selli, M. Baldoni, A. Sgamellotti, F. Mercuri, Redox-switchable devices based on functionalized graphene nanoribbons, Nanoscale 4 (2012) 1350–1354.

[15] A. Lorenzoni, M. Muccini, F. Mercuri, Correlation between gate-dielectric morphology at the nanoscale and charge transport properties in organic field-effect transistors, RSC advances 5 (2015) 11797–11805.

[16] M. Seri, F. Mercuri, G. Ruani, Y. Feng, M. Li, Z.-X. Xu, M. Muccini, Toward real setting applications of organic and perovskite solar cells: A comparative review, Energy Technology 9 (2021) 2000901.

[17] A. Lorenzoni, A. M. Conte, A. Pecchia, F. Mercuri, Nanoscale morphology and electronic coupling at the interface between indium tin oxide and organic molecular materials, Nanoscale 10 (2018) 9376–9385. doi:10.1039/c8nr02341g.

[18] A. Lorenzoni, M. Muccini, F. Mercuri, A computational predictive approach for controlling the morphology of functional molecular aggregates on substrates, Advanced Theory and Simulations 2 (2019) 1900156.

[19] Y. Feng, Q. Hu, E. Rezaee, M. Li, Z.-X. Xu, A. Lorenzoni, F. Mercuri, M. Muccini, High-performance and stable perovskite solar cells based on dopant-free arylamine-substituted copper(ii) phthalocyanine hole-transporting materials, Advanced Energy Materials 9 (2019) 1901019. doi:https://doi.org/10.1002/aenm.201901019.

[20] W. Gao, The chemistry of graphene oxide, Graphene oxide: reduction recipes, spectroscopy, and applications (2015) 61–95.

[21] M. Baldoni, F. Mercuri, Evidence of benzenoid domains in nanographenes, Physical Chemistry Chemical Physics 17 (2015) 2088–2093.

[22] A. K. Geim, Graphene: status and prospects, science 324 (2009) 1530–1534.

[23] A. Barnard, M. Soumehsaraei, Benyamin, B. Sun, L. Lai, Neutral graphene oxide data set. v1. csiro. data collection, 2019. URL: https://doi.org/10.25919/5e30b44a7c948. doi:10.25919/5e30b44a7c948.

[24] G. B. Goh, C. Siegel, A. Vishnu, N. O. Hodas, N. Baker, Chemception: A deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models, 2017. arXiv:1706.06689.

[25] K. Mills, I. Tamblyn, Big graphene dataset, 2019. URL: https://doi.org/10.4224/c8sc04578j.data. doi:10.4224/c8sc04578j.data, 1 compressed .tar.gz file (3.77 GB) containing 60,744 .h5 testing files and 501,473 .h5 training files.

[26] F. L. Piane, M. Baldoni, M. Gaspari, F. Mercuri, Introducing mambo: Materials and molecules basic ontology, arXiv preprint arXiv:2111.02482 (2021).

[27] BIO-SUSHY, 2023. URL: https://www.bio-sushy.eu/, accessed: 2023.

[28] D. O. Rocha Cadena, Simulación y fabricación de celdas solares de perovskita híbrida con arquitectura invertida (pin), Master's thesis, Tesis (MC)–Centro de Investigación y de Estudios Avanzados del IPN …, 2023.

[29] T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A. G. Ricciardulli, A. Vijayan, et al., An open-access database and analysis tool for perovskite solar cells based on the fair data principles, Nature Energy 7 (2022) 107–115.

[30] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner, Microscopic traffic simulation using sumo, in: 2018 21st international conference on intelligent transportation systems (ITSC), IEEE, 2018, pp. 2575–2582.