EMMC Roadmap

Digital Transformation of Materials Science

EMMC ASBL

The European Materials Modelling Council

Responsible for content:

EMMC ASBL

Avenue Louise 54, 1050 Brussels, Belgium
CBE no.: 0731 621 312

EMMC Roadmap on the Digital Transformation of Materials Science

Contents

Introduction

Materials Science has relied on materials information exploration, modelling and simulation for more than 30 years. Early examples include the Cambridge Structural Database[1], available in electronic form since the 1970s, the advent of software companies with materials modelling packages such as Molecular Simulations in the 1980s[2], as well as materials information technology (Granta Design[3], founded in 1994, now part of Ansys). Furthermore, Machine Learning (ML) technologies have been employed in materials design and development for many years, see for example the QSPR approaches[4].

With the growth of digital technologies there has been a confluence of many approaches that combine to drive a data- and model-based design, development and engineering of materials. Models and data combine for example in digital twins of processes and products. As the amount of data (produced by experiments as well as modelling), and the possibilities to exploit data have grown, so has the need for better managing and organising data. While purely data science based approaches such as data lakes etc. have enabled some progress regarding the issue of data access and data siloes, it has also become clear that the materials science field itself and its knowledge of the domain is called upon to improve the way data is organised. While other fields, such as biology[5], chemistry[6],[7] and engineering[8], have advanced their metadata, terminologies, taxonomies and ontologies to improve the FAIRness[9] of data, a coherent approach for materials sciences remains somewhat lacking[10].

[1] https://www.ccdc.cam.ac.uk/solutions/software/csd/

[2] https://www.researchgate.net/publication/280934406_A_Scrolling_History_of_Computational_Chemistry#fullTextFileContent

[3] http://www.eng.cam.ac.uk/news/department-spin-out-granta-design-acquired-ansys

[4] https://www.schrodinger.com/science-articles/machine-learning-qspr-materials

[5] https://geneontology.org/

[6] https://goldbook.iupac.org/

[7] Strömert, Philip, Johannes Hunold, André Castro, Steffen Neumann, and Oliver Koepler. 2022. "Ontologies4Chem: The Landscape of Ontologies in Chemistry." Pure and Applied Chemistry 94 (6): 605–22. https://doi.org/10.1515/pac-2021-2007.

[8] https://en.wikipedia.org/wiki/ISO_15926

[9] https://www.go-fair.org/fair-principles/

[10] De Baas, Anne, Pierluigi Del Nostro, Jesper Friis, Emanuele Ghedini, Gerhard Goldbeck, Ilaria Maria Paponetti, Andrea Pozzi, et al. 2023. "Review and Alignment of Domain-Level Ontologies for Materials Science." IEEE Access 11: 120372–401. https://doi.org/10.1109/ACCESS.2023.3327725.

The importance of the availability, transparency and access to data as key factors for reaching the objectives of Materials Science has been widely recognised. There are significant initiatives at national, European and international level such as:

- **Germany:** Platform Material-Digital PMD [11], NFDI [12] including e.g. NFDI-Mat-Werk [13], NFDI4Ing[14] and NFDI4Chem[15], and FAIR-mat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids)[16]
- **France:** DIADEM – Discovery Acceleration for the Deployment of Emerging Materials[17]
- **At European level:** the European Materials Modelling Council (EMMC), European Materials Characterisation Council (EMCC), and their related projects[18,19]. Also, EOSC and its federated FAIR data spaces concept and interoperability framework[20] are important in this context.
- **At international level:** Research Data Alliance (RDA)[21] and MARDA Alliance[22]

EMMC Vision for the Digitalisation of Materials Science

The vision of the EMMC for the digitalisation of Materials Science is:

- **Data that is well documented** in a harmonised way will become easier to find, more interoperable and better reusable. This facilitates the integration of data from a wide range of sources, as well as **interoperability of modelling and characterisation methods**.
- Accessible and more widely available **capacities and capabilities in generating, combining and and exploiting materials data**: new data and insights are generated by advanced modelling and characterisation, combined with safe and sustainable-by-design principles, and data exploitation is enhanced by with Machine Learning and AI.
- **A Common Digital Materials Ecosystem with federated data repositories**, trusted data access and exchange and a well-managed materials data infrastructure at its core will bring benefits

---

[11] Platform MaterialDigital PMD: https://www.materialdigital.de/
[12] NFDI, The German National Research Data Infrastructure https://www.nfdi.de/?lang=en
[13] NFDI-MatWerk Nationale Forschungsdateninfrastruktur für Materialwissenschaft & Werkstofftechnik; https://nfdi-matwerk.de/
[14] https://nfdi4ing.de/
[15] NFDI4Chem, Chemistry Consortium in the NFDI https://www.nfdi4chem.de/
[16] FAIRmat FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids; https://www.fair-di.eu/fairmat/fairmat_/consortium
[17] DIADEM: an exploratory Priority Research Programme and Infrastructure linking materials and AI; https://www.cnrs.fr/en/diadem-exploratory-priority-research-programme-and-infrastructure-linking-materials-and-ai
[18] https://emmc.eu/emmc-related-projects/
[19] http://characterisation.eu/projects/
[20] https://www.eosc-pillar.eu/establishing-fair-data-services , https://eosc-portal.eu/eosc-interoperability-framework
[21] https://www.rd-alliance.org/
[22] https://www.marda-alliance.org/

for all actors along the materials value chains. Such an ecosystem will ensure interoperability between all 'horizontal enabler' technologies, and enable integration of the information rendered in order to accelerate the design and development of advanced materials as well as ensure materials safety and sustainability throughout their life cycle.

To work towards this vision, **efficient pathways** for generating, documenting, managing and utilising relevant data must be created and managed, as shown in Figure 1 and 2

Figure 1: Digitalisation and Data Pathways

Data-driven activities accompanying these materials and digital capacities and capabilities form an iterative cycle as shown in Figure 2. Pursuing these activities iteratively and continuously will form a virtuous cycle for advanced materials development, which can create synergistic effects along and across value chains.

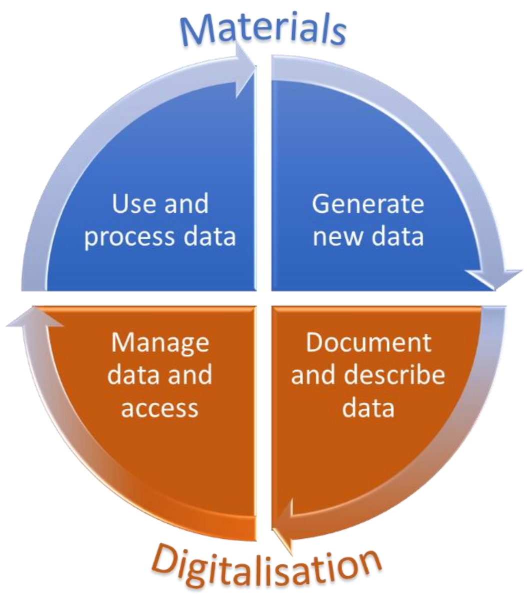

Figure 2: The data life-cycle of advanced materials

Initial work has been done in EU and national projects but optimisation and harmonisation of throughout the value chain is required.

Expert discussions have led to the following priorities and proposed actions to move towards the above vision.

Data documentation and interoperability

Documentation of data and knowledge for FAIRness

There is a need for increased FAIRness (findability, accessibility, interoperability and re-use) of data. This can be supported by semantic technologies based on a common (standardised) language (ontology) for data documentation, data exchange, knowledge extraction and management in materials science.

Information technologies (logic) can use these semantics to provide reasoning and recommendations (knowledge).

Semantic technologies rigorously represent the meaning of concepts providing an unambiguous interpretation (data documentation) of standards which will support the enforcement of standards. A specific standard terminology facilitates the implementation of certified procedures and standard measurements. Standard terminology also delivers criteria for search tools.

For data to be useful and accessible for all stakeholders, ontologies will thus need to be established that cover the materials value chain.

Initial steps in terminology, classification and data documentation for multi-perspective materials modelling and characterisation workflows have been done, establishing the now widely accepted data documentation systems MODA[23] and CHADA[24], respectively. Europe is leading the way in ontology-based data documentation of materials and manufacturing due to the development of the EMMO ontology[25], overseen by EMMC. The EMMO is an ontology for the applied sciences and can document the typical multi-perspective study of materials, dealing with all aspects of physical matter, from quantum systems to continuum materials, with data, properties, physical quantities, models, processes etc.

However, more work needs to be done to lead to a harmonised documentation of data and transfer of knowledge in a form understandable to humans and machines. This is due to the increasing complexity and diversity of data needed for the development of advanced materials integrating safe-and-sustainable-by-design aspects of extending lifetimes and increasing circular material flows, minimising the environmental impact and improving health and increasing the diversity of involved stakeholders. Novel materials require semantically documented databases containing both empirical and simulated data in multidimensional spaces beyond the configurations (e.g. chemical compositions of metal alloys) that are currently used actively.

[23] Models and Data Framework; https://modeling-languages.com/a-hitchhikers-guide-to-model-driven-engineering-for-data-centric-systems/; https://emmc.info/moda/
[24] Creation of novel metadata structure (CHADA) for collection data and information related to characterisation methodologies; https://cordis.europa.eu/project/id/760827; https://zenodo.org/record/2636609#.Yqgec3ZIDIU
[25] European Materials Modelling Ontology: https://github.com/emmo-repo/EMMO

Research shows that complexity (knowledge documentation) still is a primary hindrance in predictive accuracy of novel materials.

It must hence be a priority to continue working towards a common (standardised) language (concepts and vocabulary) for data documentation in the materials sciences that supports data exchange and interoperability, and enables communication along the value chain and across industrial fields.

Strategic Activities

a) **Develop a harmonised system of taxonomies and ontologies** supporting all relevant types of materials, development stages and circularity issues in the different industrial fields

Activities pertinent to this priority include the following examples:

- Applying and further building the OntoCommons[26] EcoSystem (OCES) of methods and tools for the development and governance of materials ontologies ready for addressing knowledge challenges in industrial fields.
- Develop a system of interoperable middle- and domain-level ontologies based on EMMO, that can support a wide range of digitalisation and automation applications in the advanced materials discovery and development cycle, including materials safety and sustainability. Requirements will be user case driven, uncovering synergies to create the harmonised and interoperable data and knowledge documentation with constant feedback from the stakeholders.

b) Establish **meaningful descriptors for materials** (defining what a material is and how it functions etc.) to derive insights from abundant data across the entire materials science field and across material value chain. These activities will need to involve all stakeholders.

Activities pertinent to this priority include the following examples:

- Develop fast and impactful indicators of materials behaviour and properties based on the descriptors, a.o. utilising large data sets.
- Support the development of materials identification e.g. to support registration, traceability and recycling/re-use

c) Standardisation of digital documentation **of safe and sustainable-by-design chemicals and materials.** That is for safety and Life Cycle Assessments and circularity.

Activities pertinent to this priority include the following examples:

- Development of standardised digital documentation of the safety and sustainability of materials, certification schemes, ecolabels (including safety and sustainability, risk & lifecycle assessment). This documentation system should inform all stakeholders including consumers with materials and substance data in a standardised way.

d) Standardisation of digital documentation for all **materials data generation technologies and methods** (including computational modelling, materials synthesis and characterisation methods and autonomous robotised synthesis)

---

[26] https://ontocommons.eu/

Activities pertinent to this priority include the following examples:

- Establish standardised digital documentation (machine readable) for **new data generation methodologies** encompassing all data and meta-data required for understandability and reproducibility of the methods and their results. The documentation should include study design, protocols, SOPs, quality control, and confidence levels.

e) Establish standardised digital data documentation that represents material data on **validation/verification and quality assessment**

Harmonise and integrate materials multi-technique (e.g. modelling and characterisation) techniques and workflows

Develop modelling and its linkage to characterisation, production and testing technologies in order to get a consistent data space.

The advanced materials lifecycle includes development, production, use, repair, recycle, and reuse supported by modelling, characterisation and testing, advanced monitoring and sensing, but also robotics and automation. These sprout data that is to be linked.

Facing the complexity of materials, it is commonly agreed that advances in simulation methodologies can help in predicting efficiently the effective physical properties of multiphase materials with complex internal microstructures. Analysing characterisation images elucidate the chemistry and physics of complex materials. New machine learning (AI) methods have been shown very efficient in predicting the properties of materials by combining large amounts of simulation and characterisation data and combining high amounts of low-fidelity data with low amount of high-fidelity data. However the quality of the data is central for the success of these approaches and knowledge could be fed into these approaches by semantic technologies.

Some key gaps need to be addressed such as the lack of optimised and harmonised data categorisation and description; the lack of standardisation of methods and machine-readable, standardised output from devices; the difficulties to acquire multi-source, multi-fidelity data regarding uncertainty, cost, time, data volume, phase space etc.

The workflow from raw data to final result needs to be digitalised. This may involve data from modelling, characterisation and robotics. It should lead to workflow optimisation.

Strategic Activities

Activities pertinent to this priority include the following examples:

a) Assess and harmonize current protocols and standard methods to enable the establishment of multiscale and multitool by utilizing an interoperable data structure.

b) There is a need to build digital workflows integrating the tools

c) Develop the interface with the cloud, to store and analyse the information, independently of the physical testing.

d) Integrate multi-techniques that are specifically optimised towards the innovation markets to maximise the potential application and value of data.

e) Develop workflows able to generate multi-source, multi-fidelity data depending on need e.g. regarding uncertainty, cost, time, data volume, phase space etc.

f) Develop methods for cross-validation for high-throughput and high-speed metrology by means of different characterisation methods in order to establish standardised outcomes and methods usable to industry

g) Integrate multi-scale computational modelling, materials synthesis and characterisation methods and autonomous robotised synthesis.

h) Combine novel characterisation and modelling or virtual and physical tracking and monitoring to support the establishment of efficient materials management strategies.

i) Linking simulations with in-process data collection from e.g. autonomous robotics platforms and fabrication technologies

Develop digital strategies combining semantics and AIs

Strategic activities

a) Validate heterogeneous data processing and data correlation tools to transform the data into new, experimentally testable hypotheses that will suggest new directions for materials development.

b) Develop specific tools to improve the management of the heterogeneous experimental, simulation and manufacturing data to be used also for Digital Twins of processes. Develop guidelines for harmonising digital twin strategies across Europe

c) Develop methodologies for processing, and transforming massive data fluxes including semantic training sets for AI.

d) Develop innovative digital strategies to support the mining of materials information during service life until EoL and closing the loop, and to support circularity

e) Enable improved decision making by processing data from multiple sources into information/knowledge of added value, using semantic technologies, Machine Learning/AI and digital twin strategies.

f) Provide tools for automation.

g) Create competitive data and knowledge exploration systems (such as knowledge graphs and exploratory search systems) specific for the common materials ontologies, enabling all stakeholders to explore and interrogate knowledge about materials along and across value chains and markets.

Capacities and Capabilities

Advancing modelling and simulation development

This includes physics-based methods from electronic to continuum, multiscale and multi-physics as well as data-based and AI/ML methods. There is a need for standards for data generation: quality, new methods standards, governance and standards/benchmarks as well as VVUQ for materials models.

Strategic Activities pertinent to this priority include the following examples:

a) Development of models to cover the needs of the innovation markets, e.g. to handle bio-based materials.

b) Development of materials modelling methods and software that is user-friendly, robust and validated.

c) Development of modelling methods that are interoperable and easy to integrate in digital workflows in order to allow automatic and/or high-throughput calculations.

Education on modelling and digitalisation

Strategic Activities

a) Coordinate the development of education and training of experts with semantic, conceptualisation, and domain knowledge.

b) Training on data documentation regarding characterization and modelling of advanced materials is essential for bringing methods and related data closer to industry needs.

c) Develop the role of Knowledge Management Translator and Benefits Advisor fostered by OntoCommons, who will enable a broader community to participate in the materials digitalisation and disseminate the advantages to make people adopt.

d) Develop professional development in semantic technologies for materials domain experts.

e) Assist participation in the common materials dataspace via s standard way to communicate materials to consumers should be developed.

f) Develop and disseminate validated methods for materials data exploration, query, evaluation

Benefits Advisors

Developing roles similar to the Materials Modelling Translator and the Knowledge Management Translator is key to delivering on the opportunities of the digital transformation in materials science. In particular, the skills need to be developed that provide value assessments and advice on how benefits can be achieved by business transformations are required.

Strategic Activities

a) Test and continue to develop a Benefits Tool developed in a Task Group of the EMMC addressing business, economic and technical benefits at different levels of a company.

b) Develop a similar benefits tool for the application of semantic technologies in materials industries.

Common Materials Ecosystem

There is a need to develop a common materials ecosystem based on existing marketplaces linking distributed data repositories with trusted management, data access and exchange methodologies. This ecosystem should provide reliable and easy access to and exchange of generated data/information/knowledge for all stakeholders. The ecosystem should networking of digital resources. While developing new data generation methods the interface to other materials science activities should be developed. This includes interfacing with data resources and their semantic management. The goal is to arrive at optimised networks that can be used to develop new approaches and design tools for resources optimisation, customisation, managing circularity, improved performance, and durability, which all require specific data and cross domain data management.

Development and market adoption of new advanced materials will require management of and access to a diverse set of information related to characterisation, functionality, efficiency, reliability, safety, life cycle assessment and circularity criteria.

It should also be stressed that this management is not only for data in its narrow sense but applies to all research output needed to understand and potentially even reproduce the data (e.g. study design, measurement principle, experimental setup, protocols, SOPs, quality control, confidence levels, processing and analysis workflows, software). Such data is created by many different players ranging from academic institutions, GOs and NGOs to industry, SMEs and consultancy firms.

Strategic Activities

Activities pertinent to this priority include the following examples:

This market space should take recent EU project developments into account

a) A data space is to be developed that should consist of federated data resources connected via the semantic data documentation system developed above

- Elaborate standardised interfaces between all horizontal enablers based on the common documentation systems.
- Elaborate the use of blockchain technology which has great potential to provide transparency and communication in global value chains. The protocol enables trusted secure data exchange in fragmented supply chains while protecting a company’s privacy and sensitive information. Achieving a standard for traceability to origin would enable the proof of origin of materials, therefore fostering recycling practices.

- Support activities that process existing data to higher standards regarding validation, accuracy and confidence levels, in order to increase quality and trust in data. Well documented and curated data are the main resource for artificial intelligence models to establish meaningful and dependable correlations of manufacturing/process/outcome regarding materials characterization and modelling.

b) Technical Management of the common material data space based on federated data management. To bring this wide range of data into the common digital ecosystem serving various materials innovation market there need to be a flexible and federated data space with a management infrastructure coordinating and addressing the requirements of all stakeholders with respect to openness vs. privacy/IPR protection, security, transparency and trustworthiness. Develop the practical management of the market space. Establish materials information management and catalogues which should also include licence handling. Changes in data sources need to be registered (data management).

c) Develop the governance of the materials data space respecting EU rules and values on Data Governance, control and security, openness, interconnection, intellectual property rights, and interoperability. The European data strategy including the Data Governance Act will play an important role. Confidentiality wished for the data owned by different owners should be catered for.

d) Establish a common agreement to follow the FAIR principles, applied to data as well as protocols, analysis workflows and software, and comply to common data management and sharing standards. Define data provenance responsibility and open accessibility, set a framework for accessibility. Promote the use of a data management plan based on the harmonised data documentation system.

e) Develop new repositories in areas where needed and transform existing data into reliable, curated materials data repositories for critical aspects of the innovation markets and their value chains. The databases should be documented with the semantic system which includes storage of metadata in a searchable repository

- Develop new data repositories that combine different aspects of materials design, development and life-cycle, such as materials manufacturing data from highly digitalised processes, with materials properties and durability, from testing and multiscale simulation data, under a governance that will allow openness and confidentiality. This should include documentation of failures.

- Transform existing data by documenting it with meta data using the common semantic data documentation system. This will ensure that the documentation is prepared in a structured way using a data schema, which can be communicated as part of the data transfer, so that it can be integrated and combined with other data using automated (meta)data schema mapping.

Conclusions

This roadmap section identifies the key challenges and provides priorities on materials innovation driven by digitalisation. Pushing materials frontiers will require collective efforts of a diverse community of materials science, data science and semantic technologies expertise and innovations from a broad spectrum of disciplines.

Strategic directions and actions for data documentation and interoperability of data generation methods, improved data exploitation, capabilities and capacities in advanced modelling and simulation, human resources and the establishment of a digital materials commons ecosystem have been put forward.

The ecosystem will unite and orchestrate digital and materials capacities and competences to accelerate and optimise all aspects of materials design, development, use and re-use. Capacities include data documentation based on ontologies, materials modelling and characterisation, and robotics and Machine Learning/AI. This common materials ecosystem could be served with an open two-sided market place platform where all providers and users can meet. It will assist companies to participate in this emerging ecosystem and bring their internal systems fully up to date with forward-looking technology platforms that are open and connected.

Acknowledgements

EMMC would like to thank its members and discussions with collaborators in the Advanced Materials Initiative AMI2030[27] for their contributions to this roadmap.

Editors: Gerhard Goldbeck and Anne de Baas (EMMC, Goldbeck Consulting Ltd)

The Editors acknowledge support by the European Union Horizon 2020 research and innovation programme for project OntoTrans, Grant Agreement No 862136.

[27] https://www.ami2030.eu/