

NANOINFORMATIX PLATFORM: DATA AND MODELLING INFRASTRUCTURE FOR SAFETY ASSESSMENT AND SAFE BY DESIGN OF ENGINEERED NANOMATERIALS USED IN CONSUMER PRODUCTS AND MEDICAL APPLICATIONS

Gianpietro Basei¹, Alex Zabeo¹, Nicola Toschi², Antonio Pietroiusti², Lisa Pizzol¹, Matteo Carisi¹, Giacomo Chiarot¹, Danail Hristozov³

¹Greendecision Srl, Venezia, Italy

²Univeristà degli studi di Roma Tor Vergata, Roma, Italy

³East European Research and Innovation Enterprise, Sofia, Bulgaria

Abstract

In the NanoInformaTIX project we are developing a sustainable data management and multi-scale modelling framework to facilitate the (eco)toxicity and exposure assessment of engineered nanomaterials (ENMs) for the purposes of more cost-effective risk analysis and safety by design. This framework will be implemented as a web-based platform specifically tailored to the needs of stakeholders from industry, regulators and academia.

The platform will be implemented as a digital EU hub connecting relevant databases to nanoinformatics models covering Materials, Exposure, Physiologically-Based Pharmacokinetics (PBPK), Quantitative-Structure-Activity Relationships (QSAR) and System Biology to support in vitro-in vivo extrapolations (IVIVE), grouping and read-across.

The databases and models will be linked to the platform by means of a common Application Programming Interface (API) to enable real-time updating of the models by their developers. To establish interoperability between models and/or data repositories we adopted the Modelling Data Elements (MODA) scheme of the European Materials Modelling Council. This will enable the user to design chains of data sources and models corresponding to specific analyses (e.g. prediction of genotoxicity, dose-response assessment, IVIVE). The model inputs and outputs will be clearly documented, while the results of the analyses will be presented as both dynamic charts and textual reports.

To obtain optimal confidence in the use of the platform, the models, their applicability domains, and the steps to create the model chains will be carefully documented according to established standards and Good Practices.

Keywords: *engineered nanomaterials, risk assessment, data management, in silico modelling, computational infrastructure*

1. INTRODUCTION

Experimental studies, over the last decades, have generated a considerable amount of data on the potential (eco)toxicity of engineered nanomaterials (ENM) and their physicochemical characteristics [1]. Still, predicting the risk posed by a product based on its physicochemical properties is very challenging, as a product may have a complex structure that usually undergoes intricate transformations during its lifecycle. Research into the underlying links between adverse effects and physicochemical structure have intensified over the last ten years, with the concept of Adverse Outcome Pathways becoming part of the OECD initiative for Integrated Approaches to Testing and Assessment. Evidently, modelling will be central to future integrated concepts; at the very least to depict the current theories on key events, pathways and parameterisation but very likely as a powerful tool for in silico safety testing, and design-for-safety. Then, modelling would be a major driver for cost effectiveness in product development and market authorisation, creating real innovation potential. We hypothesise that, using most modern informatics methods [2], it is feasible to establish a framework for the *in silico* safety assessment of ENM in each stage of an ENM's life cycle.

To achieve this, In the European H2020 research project NanoInformaTIX project we are developing a sustainable data management and multi-scale modelling framework to facilitate the (eco)toxicity and exposure assessment of ENMs for the purposes of more cost-effective risk analysis and safety by design. This framework will be implemented as a web-based platform specifically tailored to the needs of stakeholders from industry, regulators and academia. To effectively develop and implement this sustainable data and modelling platform (SDMP), we have set the following objectives:

- Establish the SDMP as a digital European hub to ease the access of industry, regulators and academia to predictive modelling and data repositories, and to facilitate the training and validation of models based on comprehensive curated datasets
- Connect the systems of databases and modelling tools selected and tested/validated in the project to the SDMP by means of a common Application Programming Interface (API)
- Engage in stakeholder consultations in order to align the SDMP objectives and functionalities with the actual needs of its potential end-users
- Test the sensitivity and performance of the SDMP (the focus in WP7 is not on the individual models, which will be tested in WP6, but on the platform as a whole)
- Document the SDMP and its underlying database and models according to Good Modelling and Programming Practices based on established standards and Standard Operating Procedures (SOPs), and develop detailed User Guidance

2. MATERIALS AND METHODS

The NanoInformaTIX SDMP will be implemented as a digital EU hub connecting a state of the art data management system to suite of validated models: Materials modelling, Exposure modelling, Bio-distribution modelling: Physiologically-Based Pharmacokinetics (PBPK), Dose-Response modelling: Quantitative-Structure-Activity Relationships (QSAR) and Systems Biology modelling to support the prediction of the potential (eco) toxicity of ENMs during each stage of their lifecycle for both regulatory risk assessment and safe by design purposes.

Definition of the common API

The databases and models will be linked to the platform by means of a common Application Programming Interface (API). We will define a common RESTful API (Representational State Transfer Application Programming Interface). The API will consist of the definition of the minimum set of input (alongside with data type and domain) and the expected outputs for each model. APIs will be organized hierarchically as a) generic, b) model category related and c) model specific. The common API will be written specifically to be easily extended by adding new models or extending existing ones by defining additional specific input and output for specific model categories (e.g. (Q)SARs, PBPK). This will allow modelers to keep their models updated, and to include new models in the platform, thus making the platform open and transparently upgradable.

MODA-based interoperability of data and models

To establish interoperability between models and/or data repositories we are adopting the Modelling Data Elements (MODA) scheme of the European Materials Modelling Council. In the MODA workflow, the system is able to process model output to feed it into the next model of the chain (cf. Figure below) by means of Extract Transform Load (ETL) standard operations. Hence, each model in the chain will be able to perform its analysis by using data from data repositories, parameters computed by previous models in the chain, as well as additional user inputs. Each chain will correspond to a specific operation needed for risk assessment (e.g. in vitro-in vivo extrapolation (IVIVE), dose-response assessment, grouping, read-across) or for prediction of specific properties or endpoints needed to develop strategies for safer design of quality products. The interfaces of the platform will be designed to allow such operations to be easily selected by the user as each selection will generate chain(s) of data

sources and modelling tools specific to the desired analysis. The results will be presented as dynamic charts and textual reports.

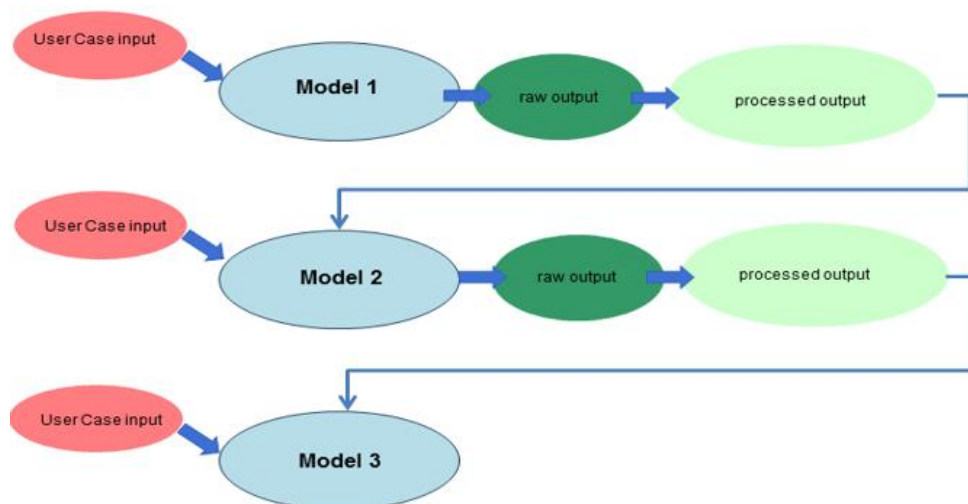


Fig. 1. A schematic of the MODA workflow. The system is able to process model output to feed it into the next model of the chain.

This MODA-based system will enable the user to design chains of data sources and models corresponding to specific analyses (e.g. prediction of genotoxicity, dose-response assessment, IVIVE) of importance for risk assessment and SbD. Once the chains are defined, the user will be guided in selecting descriptors/endpoints and providing model-specific inputs. The software will provide detailed information on which models in the chain have already produced outputs, which ones are ready to be run, and which are waiting for user-inserted data or inputs from previous models in the chain. The model inputs and outputs will be clearly documented, while the results of the analyses will be presented as both dynamic charts and textual reports.

The SDMP will address the OECD principles for the validation of (Q)SARs. Specifically, it will allow the users to validate their models, highlight their applicability domains and provide information on goodness-of-fit statistics. Moreover, it will be possible to assess the robustness of the models and to study the uncertainty in the obtained results. Moreover, the platform will allow to perform model-specific and global (i.e. involving the whole chain of models) sensitivity analyses.

Software implementation of the SDMP

The SDMP web portal will be JavaScript-based and developed as a Meteor application. Meteor is a complete framework which allows for a fully JavaScript programming environment with a Node.js server and Mongo.db No-SQL database. Meteor implements a novel Distributed Data Protocol which supplies wireframing for fast and reliable optimistic User Interface and live updates. The dynamic user-friendly charts for model output display to the users will be provided thanks to the Highchart.js package. The platform will be able to scale and run different models concurrently, like it was implemented by GD in the SUNDS system.

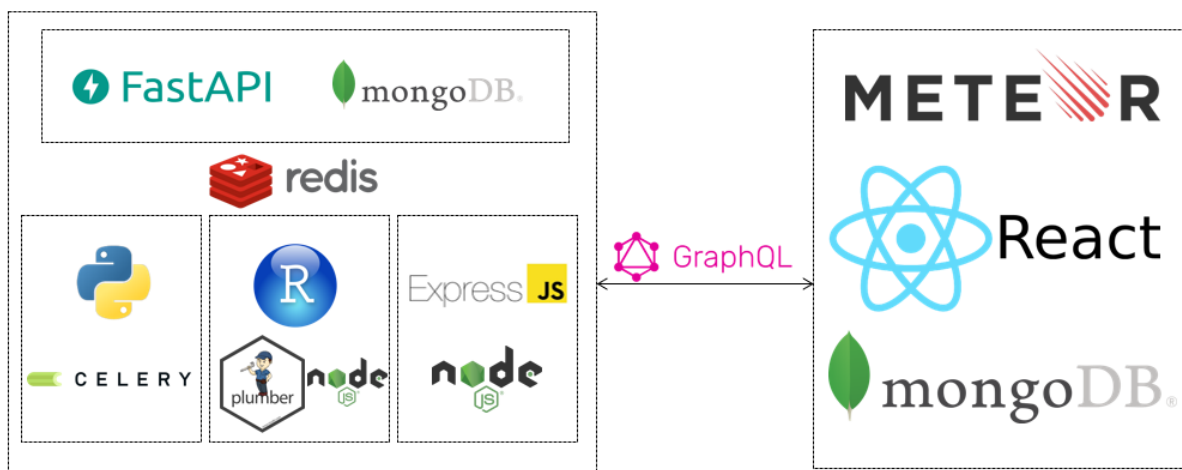


Fig. 2. Schematic overview of the technical implementation of the current version of the platform. The frontend is based on Meteor and React, while the backend is based on FastAPI, and currently supports executing models programmed in R, Python or JavaScript. Data is stored in MongoDB.

3. RESULTS AND DISCUSSION

We have automatically built an API starting from associated MODA files and created an additional Manifest File that contains the technical details needed for the models to be run in the platform. We also developed the first prototype of the SDMP platform as a web-based application and started inserting first examples of chains of models for human health and environmental risk assessment. We also made some first steps in the demonstration of the SDMP to stakeholders and developed the approach for its performance testing in case studies.

Demonstration and testing of the SDMP

To obtain optimal confidence in the use of the platform, the models, their inputs and outputs, their applicability domains, and the steps to create the model chains have been demonstrated to stakeholders to elicit their constructive feedback as a basis for further developing the platform. In addition, we will carefully document the system and its underlying models according to established standards and Good Practices.

Specifically, we 1) have demonstrated the SDMP to stakeholders in order to elicit their constructive feedback on its objectives, user interfaces and functionality, and once the prototype of the system is ready, we 2) have developed the methodology for its sensitivity and performance testing, and 3) have planned the development of detailed user Guidance and documentation for transparency.

Demonstration of the SDMP to stakeholders

The SDMP has been demonstrated to relevant stakeholders in a workshop and a training school, so that their feedback is collected and used to improve the platform's performance. Indeed, from the start of the project we have engaged with data scientists and modelers as well as with regulators, industries, consultancy companies, standardization bodies, and the civil society in a continuous dialogue as we develop the portal.

Our stakeholders are also the existing software frameworks and tools for risk assessment and management of ENMs and nano-enabled products (e.g. SUNDS, GUIDEnano). These web-based systems can benefit greatly from connecting to the databases and the models in the SDMP, which can enable them to automatically acquire curated data and perform different analyses (e.g. predictions of properties and effects as part of SbD strategies; extrapolations, grouping and read across for risk assessment), which would substantially empower these tools and extend their applicability domains. Therefore, we have worked with the developers of these systems to ensure that they can seamlessly

connect to the SDMP using the common API and that our platform can effectively enable them to assess the risks from ENMs and support risk management decision making.

In demonstrating the SDMP specific consideration has been given to: 1) the (further) needs and perspectives of the stakeholders, 2) the added value of the platform towards their objectives; and more generally 3) the practical usability of the SDMP for informing regulatory risk assessment and safer design of quality products. Continuous dialogue with the stakeholders is essential for ensuring credibility, and therefore the future of the platform. Our targeted efforts to align it with end-user needs, to raise awareness, provide training and develop a business plan to ensure that the SDMP has a legacy beyond the duration of the project.

Sensitivity and performance testing of the SDMP

The SDMP will be subjected to sensitivity analysis and performance testing. The tests will cover capabilities such as selection of models and input data sources, chaining them for specific analyses (e.g. IVIVE, dose-response, grouping, read-across), successfully processing the outputs of certain models to use as inputs to others, graphically representing the results and the uncertainty in the analyses.

To perform the sensitivity analysis, the one-factor-at-a-time method will be applied by running the selected model chains multiple times and computing the results by changing input parameter values in each trial. These new values will be drawn according to Monte Carlo sampling which implies the use of a specific probability distribution for each input parameter, but where a distribution is not available, uniform or Gaussian distribution will be used. In addition, modelling errors, possible propagations of errors from models providing input to other models, and the estimation of the final uncertainties in the outputs of the SDMP will be considered. The modelling errors will be available from the detailed regression statistics during the model development. The quantification of propagated errors will take into account the combined uncertainties of the models. The relevant algorithms will be included in the SDMP, so that the uncertainty analysis can be applied automatically when new models are uploaded to the platform after the end of the project.

SDMP User Guidance and documentation for transparency

To facilitate the use of the SDMP, a detailed user manual will be developed, which will be tailored to the specific needs of our stakeholders. This guidance will contain details for user management covering both the use of the SDMP as a whole and the independent application of each database or model. In addition, we will prepare tutorials based on training datasets and pre-specified use cases which new users can go through online to familiarize themselves with the platform.

The models included in the platform, their inputs and outputs, applicability domains, and the steps to create the model chains will be carefully documented according to established standards, SOPs and Good Practices. Specifically, we will adopt the EMMC's recommendations on standards for modelling software development and the REACH QSAR Model Reporting Format – QMFR.

4. CONCLUSIONS

The NanoInformatIX SDMP will contribute to enhancing the innovation capacity of various nanotechnology sectors in line with the EC strategy for growth through research and innovation [3,4]. In the medium term we expect our results, to inspire new research avenues. The SDMP will open up the parameter space for the design of future materials. So far design is predominantly guided by target measures. This makes very good sense for the final product, but when used for intermediate steps this limits creativity. The SDMP would allow defining intermediate steps that can be modelled in a reliable manner. We expect this to spur the development of new products and to develop better manufacturing processes.

In the medium-to-long term, we expect the SDMP to be useful for the holistic design of nanomaterials or nano-enabled products. Our API should allow linking the SDMP to other solution used to model the manufacturing process. So, the modelling suite that helps to optimise the manufacturing of a product with specific characteristics will connect to the modelling suite that links characteristics to risks. Our

future vision is that inventors would be able to design product and production in a way that balances performance, productivity, and risks entirely *in silico*, already before starting to set up production. The trust of other modelling communities is crucial to realise this. We believe that the unmatched quality of our data combined with an unmatched selection of most advanced modelling approaches will be a crucial strength of our project, in this regard. The modelling community of researchers will greatly benefit from our ready-to-use curated datasets, spanning endpoints of regulatory importance, and from open source and/or open access modelling components, developed in collaboration with experts from the respective scientific domains. The inclusion of data quality and completeness criteria, including information of what is technically and analytically feasible by the experimental setups will be a unique asset towards increasing the trust and validity of the model results. The modelling community will also benefit from interoperability of data and modelling components, allowing dynamic retrieval and analysis of data, beyond the static datasets. This will enable researchers from industry and universities for faster and cheaper development of new products, which will open new market opportunities.

ACKNOWLEDGEMENTS

NanoInformaTIX project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 814426.

REFERENCES

1. Karcher S et al. Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations NanoImpact 85-101 2018
2. Fourches et al. A. Exploring Quantitative Nanostructure-Activity Relationships (QNAR) Modelling as a Tool for Predicting Biological Effects of Manufactured Nanoparticles. Combinatorial Chemistry and Highthroughput Screening 217-225 2011
3. The EU 2020 Strategy (COM (2010) 2020), which defines smart, sustainable and inclusive growth as the main European 2020 objective leading to jobs, growth and further commercial investment through new market opportunities and expansion of companies.
4. EU Green Paper 2011 (COM (2011) 48), which highlights research and innovation as the key drivers of European social and economic prosperity.