Scientific workshop on Grouping of Nanomaterials: NanoReg2 and GRACIOUS projects 12 – 13 September 2018, Paris, France

Data curation, model development and integration: the GRACIOUS platform

Nina Jeliazkova[1], Peter Ritchie[2], Shahzad Rashid[2], Gianpietro Basei[3], Alex Zabeo[3], Georgia Tsiliki[4], Danail Hristozov[3]

[1]Ideaconsult Ltd, Sofia, Bulgaria; [2]IOM, Edinburgh, UK;
[3]Green Decision, Venice, Italy; [4]ATHENA RIC, Athens, Greece

Objectives

Establish a data curation system

- built upon the eNanoMapper database,
- to support grouping and read-across for regulatory risk assessment, hazard classification, and industrial Stage-Gate decision making,
- to ensure that the ***in silico* models** identified and/or improved by WP2-5 can link with the data curation system to form an interoperable infrastructure.

Enable discovery of new hypotheses

- on nano-bio interactions, modes of action (MoA) and/or Adverse Outcome Pathways (AOPs) via exploratory data analysis on the entire set of available data, and
- feed the results into WPs 2-5 for developing grouping Integrated Approaches to Testing and Assessment (IATA).

Integrate the data curation and modelling tools into an open platform incorporating latest technologies

- to enable retrieval of high-quality data for the safety assessment of nanomaterials, and
- predictive *in silico* modelling as part of the IATA.

Develop a blueprint document

- as a technical guidance on programming the GRACIOUS framework and its underlying IATA into the current and future software-based nanosafety assessment systems (e.g. SUN Decision Support System, GuideNano Tool) to enable them to perform grouping, read-across and/or classification.

Expected Results

WP6 will establish a progressive data curation system seamlessly linked to *in silico* models leading to an interoperable data and modelling infrastructure. WP6 will explore existing and develop new (dis)similarity methods for NMs as well as predictive and processing models of various complexity. Key challenges that will be addressed are the variety and veracity of the data. A tentative list of expected results in these regards includes:

- a fully optimised data management and data curation system with respect to user friendliness and documentation together with much contextual, meta and other supporting data,
- a suite of *in silico* tools for grouping and read-across, specifically adjusted and designed to the needs of WP2-5 partners,
- a high level of interoperability and integration between data and the *in silico* tools.

The GRACIOUS Framework (and its IATA) will be delivered as a blueprint (i.e. design document) describing in detail both its structural and behavioural aspects (algorithms, decision trees/tables, rules), and enabling the programming of the framework in any software-based safety assessment system. Validation of the blueprint, its structures and algorithms will be implemented in two different software tools for risk assessment and management of nanomaterials (i.e. the SUN Decision Support System and Guidenano Tool).

References

Jeliazkova, Nina, et al. "The eNanoMapper database for nanomaterial safety information." *Beilstein journal of nanotechnology* 6 (2015): 1609.
Chomenidis, Charalampos, et al. "Jaqpot Quattro: a novel computational web platform for modeling and analysis in nanoinformatics." *Journal of chemical information and modeling* 57.9 (2017): 2161-2172.
Mu, Yunsong et al. "Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs." *Nanotoxicology* 10.9 (2016): 1207-1214.
Hristozov, Danail, et al. "A weight of evidence approach for hazard screening of engineered nanomaterials." *Nanotoxicology* 8.1 (2014): 72-87.
Basei, Gianpietro, et al. "Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of *in silico* tools: a critical review." *(submitted)*

Tasks and Interim Results

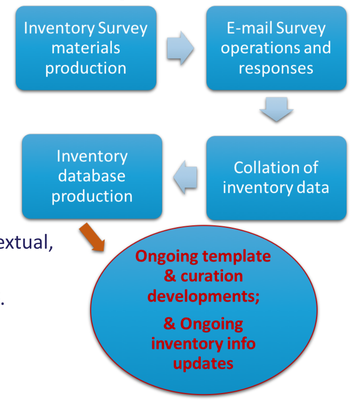
Task 6.1 Contribute to the overall GRACIOUS Framework design and refinement

An inventory of databases and *in silico* tools for the safety assessment of nanomaterials was created. The identified databases and tools were reviewed resulting in the following conclusions: (i) more efforts are needed to curate the available nanosafety datasets to enable more robust *in silico* predictions; (ii) the available models need to be validated, to avoid overfitting and to provide more reliable results; (iii) the application domains of the models are fundamental: further explore methods for read-across approaches; (iv) the uncertainty and sensitivity of *in silico* models are essential for interpreting their results; (v) advanced ML techniques should be explored (e.g. Transfer Learning) to face the limitations of currently adopted approaches.

Task 6.2, 6.3 Data curation system & new capabilities to store release and exposure data

The Data and Modelling Inventory Survey

- Part of the mandatory Data Management Plan (DMP) of the project
- Under the grouping hypotheses being addressed by GRACIOUS, identify all experiments and assays planned, with existing or new requirements for data collection templates, SOPs, and related materials, work-flows, and timescales
- Plan the GRACIOUS-eNanoMapper database development
- Obtain/curate datasets timeously, upload to the database as the project progresses
- Identify the various types of data-related analysis and modelling tools (WPs 2-5)



Results: Currently information on 49 distinct tests/assay has been obtained, together with contextual, meta and other supporting data to inform the DMP, detailed data template development and other data curation management tasks, as well as the wider modelling landscape for GRACIOUS.
Initial database online: https://search.data.enanomapper.net/gracious
(building upon NanoReg 2 curated content, including FP7 NanoReg, MARINA, NanoTest data)

Task 6.4 Interoperability and integration of modelling tools

Different levels of integration are considered given the model complexity: (i) Application Programming Interface (API) solutions; (ii) programmatic retrieval of data; (iii) data download.

Task 6.5 Exploratory modelling tools

Machine learning and statistical methodologies are considered to best match the needs of the project: i.e. estimate patterns in the data, help generate insights or new grouping hypotheses to be verified, facilitate grouping and read-across. Currently, a Subspace algorithm as a tool for read across and categorization of NMs is being employed and compared to Support Vector Machine performance.

Task 6.6 Weight of Evidence approach for hazard classification of nanomaterials

Identify hazard classes and classification criteria to develop a scientific Weight of Evidence approach for classification of nanomaterials and test it in the GRACIOUS case studies.

Task 6.7 Blueprint to enable existing software tools for risk assessment & management of nanomaterials to perform grouping and/or read-across

An object-oriented approach has been adopted in developing the blueprint document based on class-diagrams describing the rules for grouping and read-across. The class diagrams will be converted into web ontologies and integrated into the eNanoMapper ontology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 760840