

journal homepage: www.elsevier.com/locate/csbj

Review

NanoSolveIT Project: Driving nanoinformatics research to develop innovative and integrated tools for *in silico* nanosafety assessment



Antreas Afantitis^{a,*}, Georgia Melagraki^a, Panagiotis Isigonis^a, Andreas Tsoumanis^a, Dimitra Danai Varsou^a, Eugenia Valsami-Jones^b, Anastasios Papadiamantis^b, Laura-Jayne A. Ellis^b, Haralambos Sarimveis^c, Philip Doganis^c, Pantelis Karatzas^c, Periklis Tsiros^c, Irene Liampa^c, Vladimir Lobaskin^d, Dario Greco^e, Angela Serra^e, Pia Anneli Sofia Kinaret^e, Laura Aliisa Saarimäki^e, Roland Grafström^{f,g}, Pekka Kohonen^{f,g}, Penny Nymark^{f,g}, Egon Willighagen^h, Tomasz Puzyn^{i,j}, Anna Rybinska-Frycaⁱ, Alexander Lyubartsev^k, Keld Alstrup Jensen^l, Jan Gerit Brandenburg^{m,n}, Stephen Loft^o, Claus Svendsen^p, Samuel Harrison^o, Dieter Maier^q, Kaido Tamm^r, Jaak Jänes^r, Lauri Sikk^r, Maria Dusinska^s, Eleonora Longhin^s, Elise Rundén-Pran^s, Espen Mariussen^s, Naouale El Yamani^s, Wolfgang Unger^t, Jörg Radnik^t, Alexander Tropsha^u, Yoram Cohen^v, Jerzy Leszczynski^w, Christine Ogilvie Hendren^x, Mark Wiesner^x, David Winkler^{y,z,aa,bb}, Noriyuki Suzuki^{cc}, Tae Hyun Yoon^{dd,ee}, Jang-Sik Choi^{ee}, Natasha Sanabria^{ff}, Mary Gulumian^{ff,gg}, Iseult Lynch^{b,*}

^a Nanoinformatics Department, NovaMechanics Ltd, Nicosia, Cyprus^b School of Geography, Earth and Environmental Sciences, University of Birmingham, B15 2TT Birmingham, UK^c School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece^d School of Physics, University College Dublin, Belfield, Dublin 4, Ireland^e Faculty of Medicine and Health Technology, University of Tampere, FI-33014, Finland^f Misvik Biology OY, Itäinen Pitkätatu 4, 20520 Turku, Finland^g Karolinska Institute, Institute of Environmental Medicine, Nobels väg 13, 17177 Stockholm, Sweden^h Department of Bioinformatics – BiGCaT, School of Nutrition and Translational Research in Metabolism, Maastricht University, Universiteitssingel 50, 6229 ER Maastricht, the Netherlandsⁱ QSAR Lab Ltd., Aleja Grunwaldzka 190/102, 80-266 Gdansk, Poland^j University of Gdansk, Faculty of Chemistry, Wita Stwosza 63, 80-308 Gdansk, Poland^k Institutionen för material- och miljökemi, Stockholms Universitet, 106 91 Stockholm, Sweden^l The National Research Center for the Work Environment, Lersø Parkallé 105, 2100 Copenhagen, Denmark^m Interdisciplinary Center for Scientific Computing, Heidelberg University, Germanyⁿ Chief Digital Organization, Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany^o UK Centre for Ecology and Hydrology, Library Ave, Bailrigg, Lancaster LA1 4AP, UK^p UK Centre for Ecology and Hydrology, MacLean Bldg, Benson Ln, Crowmarsh Gifford, Wallingford OX10 8BB, UK^q Biomax Informatics AG, Robert-Koch-Str. 2, 82152 Planegg, Germany^r Department of Chemistry, University of Tartu, Ülikooli 18, 50090 Tartu, Estonia^s NILU-Norwegian Institute for Air Research, Instituttveien 18, 2002 Kjeller, Norway^t Federal Institute for Material Testing and Research (BAM), Unter den Eichen 44-46, 12203 Berlin, Germany^u Eschelman School of Pharmacy, University of North Carolina at Chapel Hill, 100K Beard Hall, CB# 7568, Chapel Hill, NC 27955-7568, USA^v Samueli School Of Engineering, University of California, Los Angeles, 5531 Boelter Hall, Los Angeles, CA 90095, USA^w Interdisciplinary Nanotoxicity Center, Jackson State University, 1400 J. R. Lynch Street, Jackson, MS 39217, USA^x Center for Environmental Implications of Nanotechnologies, Duke University, 121 Hudson Hall, Durham, NC 27708-0287, USA^y La Trobe Institute of Molecular Sciences, La Trobe University, Plenty Rd & Kingsbury Dr, Bundoora, VIC 3086, Australia^z Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Australia^{aa} CSIRO Data61, Clayton 3168, Australia^{bb} School of Pharmacy, University of Nottingham, Nottingham, UK^{cc} National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-0053, Japan^{dd} Department of Chemistry, College of Natural Sciences, Hanyang University, Seoul 04763, Republic of Korea^{ee} Institute of Next Generation Material Design, Hanyang University, Seoul 04763, Republic of Korea

Abbreviations: AI, Artificial Intelligence; AOPs, Adverse Outcome Pathways; API, Application Programming interface; CG, coarse-grained (model); CNTs, carbon nanotubes; FAIR, Findable Accessible Inter-operable and Re-usable; GUI, Graphical Processing Unit; HOMO-LUMO, Highest Occupied Molecular Orbital Lowest Unoccupied Molecular Orbital; IATA, Integrated Approaches to Testing and Assessment; KE, key events; MIE, molecular initiating events; ML, machine learning; MOA, mechanism (mode) of action; MWCNT, multi-walled carbon nanotubes; NMs, nanomaterials; OECD, Organisation for Economic Co-operation and Development; PC, Protein Corona; PBPK, Physiologically Based Pharmacokinetics; PChem, Physicochemical; PTGS, Predictive Toxicogenomics Space; QC, quantum-chemical; QM, quantum-mechanical; QSPR, quantitative structure-property relationship; QSAR, quantitative structure-activity relationship; RA, risk assessment; REST, Representational State Transfer; ROS, reactive oxygen species; SAR, structure-activity relationship; SMILES, Simplified Molecular Input Line Entry System; SOPs, standard operating procedures.

<https://doi.org/10.1016/j.csbj.2020.02.023>

2001-0370/© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

^{ff} National Health Laboratory Services, 1 Modderfontein Rd, Sandringham, Johannesburg 2192, South Africa

^{gg} Haematology and Molecular Medicine, University of the Witwatersrand, Johannesburg, South Africa

ARTICLE INFO

Article history:

Received 12 November 2019

Received in revised form 18 February 2020

Accepted 29 February 2020

Available online 7 March 2020

Keywords:

Nanoinformatics

Computational toxicology

Hazard assessment

Engineered nanomaterials

(quantitative) Structure–activity

relationships

Integrated approach for testing and

assessment

Safe-by-design

Machine learning

Read across

Toxicogenomics

Predictive modelling

ABSTRACT

Nanotechnology has enabled the discovery of a multitude of novel materials exhibiting unique physico-chemical (PChem) properties compared to their bulk analogues. These properties have led to a rapidly increasing range of commercial applications; this, however, may come at a cost, if an association to long-term health and environmental risks is discovered or even just perceived. Many nanomaterials (NMs) have not yet had their potential adverse biological effects fully assessed, due to costs and time constraints associated with the experimental assessment, frequently involving animals. Here, the available NM libraries are analyzed for their suitability for integration with novel nanoinformatics approaches and for the development of NM specific Integrated Approaches to Testing and Assessment (IATA) for human and environmental risk assessment, all within the NanoSolveIT cloud-platform. These established and well-characterized NM libraries (e.g. NanoMILE, NanoSolutions, NANoREG, NanoFASE, caLIBRAte, NanoTEST and the Nanomaterial Registry (>2000 NMs)) contain physicochemical characterization data as well as data for several relevant biological endpoints, assessed in part using harmonized Organisation for Economic Co-operation and Development (OECD) methods and test guidelines. Integration of such extensive NM information sources with the latest nanoinformatics methods will allow NanoSolveIT to model the relationships between NM structure (morphology), properties and their adverse effects and to predict the effects of other NMs for which less data is available. The project specifically addresses the needs of regulatory agencies and industry to effectively and rapidly evaluate the exposure, NM hazard and risk from nanomaterials and nano-enabled products, enabling implementation of computational 'safe-by-design' approaches to facilitate NM commercialization.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	585
2. Materials and methods	586
2.1. NMs datasets and knowledge infrastructures	586
2.1.1. Data generation and curation activities	586
2.1.2. Computational approaches for data gap filling and interpolation	587
2.1.3. Dedicated NMs databases organized for modelling and informatics	588
2.2. Toxicogenomics modelling (predictive models using omics data)	588
2.2.1. Deriving signatures of NM MOA to support AOP development	589
2.2.2. Deriving signatures of NM MOA that define robust predictive models based on NM bioactivity	590
2.2.3. Developing computational methods for inclusion into a robust computational platform	590
2.3. Multi-scale modelling framework for NMs property prediction	590
2.3.1. Computational NMs descriptors	590
2.3.2. Modelling (predicting) NM corona composition	591
2.3.3. Connecting NM-coronas to biological impacts/AOPs	592
2.4. Predictive nanoinformatics modelling (using artificial intelligence methodologies)	593
2.4.1. Calculated toxicity predictors for ML models	593
2.4.2. Meta models to integrate across scales	593
2.5. NM human and environmental RA	594
2.6. NM cloud platform	595
2.7. Knowledge transfer and communication with stakeholders and RA bodies	596
3. Summary and outlook	597
CRedit authorship contribution statement	597
Acknowledgments	598
Appendix A. Supplementary data	598
References	598

* Correspondence authors.

E-mail addresses: afantitis@novamechanics.com (A. Afantitis), i.lynch@bham.ac.uk (I. Lynch).

1. Introduction

Nanotechnology has increased accessibility to novel, diverse materials with dimensions of ~1–100 nm, the size range within which novel physicochemical (PChem) properties appear and transport processes occur in living systems; such materials are being used in a multitude of industrial and consumer-goods applications. The advantages of engineered nanomaterials (NMs) over similar materials in bulk are well defined; however, in most of the cases, risk assessment (RA) of the potential hazards arising from these new materials properties is incomplete or lacking. Currently, evaluation of possible NM-related risks is an expensive, slow and complicated task, usually achieved by combinations of *in vitro* and *in vivo* experiments that estimate human and environmental hazards. Clear conclusions regarding the hazards posed by NMs are often very difficult because the interpretation of experimental results is influenced by the type of procedures and protocols applied, and studies are usually limited to just a few NMs, doses and timepoints. At this stage, it is clear that proper use of existing high-efficacy occupational technical- and personal protection equipment can be used to sufficiently reduce exposure to NMs, [41,68,175] and that the effects from acute exposure to NMs at realistic doses mirror those from anthropogenic particle exposure. However, the challenges now involve assessing long-term low-level chronic exposures, and biological and ecological impacts from multi-component NMs where for example, the different components may degrade at different rates. A validated, predictive *in silico* approach that takes into account the complexity of NMs and the diverse environments in which they are deployed is essential for timely RA and continued progress in nanosafety research.

Despite the great success of computational methods to model and predict properties of conventional chemicals over many years, development of analogous quantitative models relating engineered NMs structure (morphology) to PChem properties and to toxic effects is still underdeveloped, due to:

- The relative paucity of reliable experimental data on the biological properties of NMs;
- Intrinsic complexity of NMs in particle size distribution, shape and degree of agglomeration compared to small organic molecules;
- A limited number of systematic studies on the dynamic interaction of NMs with available macromolecules when placed in a biological environment (such as serum, plasma and environmental compartments) to potentially form coronas which then become the “biologically relevant entity” seen by cells and organisms, and the role thereof.
- Systematic studies conducted to date are limited only to a few nanostructure descriptors and/or physicochemical properties namely, size, shape and surface properties [14,115].

Despite these limitations, significant progress has been made in recent years towards understanding the drivers of NMs toxicity and ecotoxicity [188,189]. Size has long been the defining feature of nanoscale materials along with surface area, but is not in itself sufficient as a predictor of toxicity [24]. Surface charge has also been found in many studies to drive toxicity, with cationic NMs being more toxic than negatively charged materials of similar composition, likely a result of strong electrostatic interactions between cationic NMs and negatively charged membranes [42,77]. Indeed, there are several models linking zeta potential and NMs toxicity in the literature [59,99,165]. The presence of crystalline order in materials is another key driver of toxicity, known since the earliest studies with quartz silica and asbestos and applies to NMs such as TiO₂ and SiO₂, where amorphous forms have low toxicity while

some ordered structures are especially toxic [4,122]. Band-gap is another common NM property linked with toxicity, as overlap of NM and cellular conductance bands facilitates transfer of electrons and oxidative stress [190]. Surface bond strain arising from high curvature and high temperature synthesis is another feature strongly correlated with toxicity, which explains differences between materials of similar composition produced by different synthesis methods (e.g. with and without high temperatures) [154,189]. A less investigated parameter that may be important for carbon-based NMs such as CNTs and graphene-like materials is chirality [38]. Many of the properties of NMs are influenced by their surroundings, so-called extrinsic properties, such as dissolution and formation of a biomolecule corona [12,89]. Binding of specific proteins that facilitate receptor mediated uptake is also a much-investigated feature of NMs, as this drives much of the subsequent signalling [74,180]. In the environment, transformations such as sulfidation, oxidation and interaction with phosphate, can change NM stability and toxicity [135,139]. Thus, it is increasingly clear that a wide range of NMs structure descriptors and properties, many of which are interlinked, are correlated with their toxicity. Modelling can elucidate the main drivers of NM toxicity via quantum mechanical and atomistic parameters that provide important insights into reactivity and mechanisms.

Evaluating the hazards of NMs requires the integration and assessment of currently quite disparate NMs characterization and toxicity data, categorization and grouping of NMs, as well as the derivation of exposure routes, forms and concentrations, and hazard threshold levels for human health and the environment. Assessing the hazards of NMs solely based on laboratory tests is time-consuming, resource intensive and constrained by ethical considerations [154]. Consequently, over the past couple of decades, computational approaches for modelling the relationships between NM structure, properties and their biological effects have become a key priority. This field has been reviewed by Winkler et al. [182], but has since advanced further. The most successful computational models, capable of predicting biological properties of NMs in diverse and complex environments, are based on the quantitative structure–activity relationship (QSAR) method. This, and related methods, use statistical and machine learning (ML) algorithms to model relationships between a materials’ structure, molecular properties, provenance and other parameters, and their biological effects. The application of these methods to diverse materials and NMs has been comprehensively reviewed by a number of authors [15,73,183], providing some very useful tools. Although data driven, these computational methods are still capable of modelling relatively small data sets [36]. Clearly, larger integrated data sets, and future as yet to be produced data, form the basis for more comprehensive models that will increase automation of experimental data analysis and interpretation. These computational approaches can rapidly fill data gaps, exploit ‘read across’ prediction of biological effects of similar materials, and classify the hazards of NMs to individual species. They are also valuable adjuncts to experimental data on the biological properties of new materials, which remains the bottleneck for reliable risk predictions [1,55,97,172]. Indeed, *in silico* models of this type are used extensively by scientists in academia and industry to reliably calculate PChem properties, and to evaluate effects on human and environmental health, ecotoxicological behaviour and fate of a broad range of chemical substances including complex materials. More importantly, integrating these data resources across disciplines, such as chemoinformatics, systems biology and ‘omics’ approaches etc., including with non-nanotechnology resources, will support multiple objectives including the reuse of existing information [60]. In addition, this deeper analysis may lead to new discoveries fuelling the innovation pipeline.

2. Materials and methods

Below we summarize the latest advances in five important fields of the nanoinformatics sector and how they can be further explored, expanded and incorporated in future generations of NM computational modelling tools and infrastructure. These include: i) dataset curation, quality assessment and knowledge infrastructures, ii) toxicogenomics modelling, iii) multi-scale modelling (physics-based and data driven), iv) predictive modelling (data driven) and v) NMs human and environmental RA. Despite the continuous advances in each sector, researchers face important challenges that need to be solved urgently. These advances are exploited within the European Commission Horizon2020 funded project [NanoSolveIT](#) to create the relevant nanoinformatics e-platform to facilitate *in silico* NMs exposure, hazard and risk assessment. To this end, each sub-chapter includes a short analysis and review of the state-of-the-art and recent bibliography, as well as a concise presentation of how the main challenges of each field are being tackled by the authors.

2.1. NMs datasets and knowledge infrastructures

Development of nanotoxicity prediction models is becoming increasingly important in risk assessment (RA) of engineered NMs but is dependent on the availability of good datasets with high data quality, completeness [92], and quantity (in terms of numbers of different NMs, but also incorporating a range of doses, timepoints, cell or organism types and end-points). While industry is responsible for the provision of data for their specific NMs for regulatory evaluation, research is required to develop the predictive models to enable reduction of the experimental data needed for regulatory RA. However, computational researchers face significant challenges in acquiring this data needed for the development of robust models. These have included the wide heterogeneity of published literature data in terms of NMs characterization, exposure and hazard data reported, the availability of the underpinning datasets in a form useful for modelling, and in regards to the variable degrees of completeness and quality of available datasets [92,157]. Therefore, in the field of nanoinformatics, there has been special emphasis on the curation and quality assessment of nanosafety data. A small number of recent projects have thus prioritized the curation of literature data and datasets generated in past projects that ended before the current (evolving) standards for data quality emerged (e.g. NanoReg2 and caLIBRAte projects curating NaNoReg data), NanoSolveIT partners curating literature publications on NMs transcriptomics etc.).

Diversity of experimental approaches in the literature data has led to many missing values (data gaps) and differing data quality (numbers of replicates, signal to noise ratio, relevance of end-points, different experimental conditions etc.), such that assessment of the quality and completeness of the collected data is a critical issue for modellers and regulators [92]. Previously, Klimisch et al. [65] proposed criteria to assess the reliability of toxicological and eco-toxicological data based on the source of the toxicity data (i.e. whether the data were produced using international standard operating procedures (SOPs) such as the aforementioned OECD test guidelines). Thereafter, Lubinski et al. [83] expanded on these criteria to include PChem properties of NMs such as size, shape, and surface charge. Recently, Ha et al. [45] and Trinh et al. [168] further extended criteria for evaluation of the quality and completeness of NM's PChem data. The quality and completeness of these data were determined using a set of rules which specifically assigned a PChem score for each attribute reported (i.e. core size, hydrodynamic size, surface charge and specific surface area). The score for each attribute was composed

of two sub-scores; one for the reliability of the data source (score range: 0–3) and another for the reliability of the measurement method (score range: 0–2). To evaluate the quality and completeness of a dataset, the average values and standard deviation of scores of all data rows in a dataset are used [168].

2.1.1. Data generation and curation activities

To support the development of better prediction models, there has been several data generation and data curation and quality assessment activities reported. For instance, Hendren et al. [50] introduced the NMs Data Curation Initiative almost 10 years ago, and explored the critical aspect of data curation within the development of informatics approaches to understand the behaviour of NMs, while Powers et al. [130] proposed a workflow for nanosafety data curation and Robinson et al. [92] discussed various issues in the evaluation of curated NM data, such as considering its completeness and quality, where the requirements for each will depend on the purpose for which the data was generated and/or will be re-used. Examples of experimentally generated and literature mined datasets suitable for modelling have also recently emerged. For example, Puzyn et al. [131] performed an experiment to determine toxicity of 17 different types of metal oxide NMs towards *E. coli* and proposed a simple equation for prediction of the effective concentration at which 50% of the organisms were killed (EC_{50}) using the enthalpy of formation of a gaseous cation. Walkey et al. [177] published experimental data and a model utilizing protein corona fingerprints to predict NMs cellular attachment, consisting of 105 surface modified gold NMs, although it turns out that the corona isolation method utilized here inadvertently removed albumin which is usually a major constituent of NM protein coronas [76,177]. Furthermore, Oh et al. [118] conducted a meta-analysis of more than 300 published articles reporting the toxicity of quantum dots and found that only a few parameters - surface properties, diameter, assay type, and exposure time, contributed significantly to their toxicity. Gernand et al. [40] collected and analyzed literature data on the rodent pulmonary toxicity of uncoated, unfunctionalized carbon nanotubes (CNTs) and proposed that the main factors driving pulmonary toxicity of CNTs were metallic impurities, CNT length, CNT diameter, and aggregate size. Melagraki et al. [97] used *in silico* methods to investigate published datasets, constructing and validating a predictive model using an organized dataset on NMs cellular uptake of 109 NPs tested in pancreatic cancer cells (PaCa2). Recently, Ha et al. [45] extracted and compiled a dataset for 26 metal oxide NMs from 216 literature articles related to the toxicity of metal oxide NMs. Trinh et al. [168] on the other hand, collected cytotoxicity data of metallic NMs, which includes PChem properties, their measurement methods (PChem attributes), *in vitro* cytotoxicity assay conditions and resultant cell viability data (Tox attributes). Each of these datasets has been, and continues to be, re-used in the development of predictive models for NM toxicity prediction.

Importantly, a number of large EC and internationally funded projects were recently completed (e.g. [29]), describing large libraries of well characterized NMs and their accompanying hazard and/or exposure datasets. Table 1 lists these datasets, which are in various stages of curation and ontological annotation for semantic mapping and database integration, by project: eNanoMapper, NanoMILE, NanoSolutions, NANoREG, NanoReg2, caLIBRAte, NanoTEST, NanoFASE, the Nanomaterials Registry, the 2 NSF-funded centers for environmental implications of NMs (CEINT and CEIN) and South Korea's S2Nano, for which several relevant PChem and biological endpoints have been assessed mostly using OECD methods.

Based predominantly on OECD documents from 2005 to 2017, Steinhauser and Sayre reviewed and summarized the key PChem properties, their preferred measurement metrics, as well as

Table 1

Datasets from various sources contributed by NanoSolveIT partners which are currently being curated and ontologically annotated by NanoSolveIT for use in modelling and federation into a knowledge commons.

Projects	Materials	Information included	Numbers
NanoMILE	Diverse NMs (i.e. ZnO, CuO, Au, Ag, CoO, SiO ₂ , BaTiO ₃ , AlOOH, Si, CeO ₂ , CuO, hydroxyapatite) with nanoinformatics and <i>in vitro</i> toxicity data	Size dependent nanodescriptors	>3000 datapoints
NanoSolutions	A panel of 30 industrial NMs each with variants of surface – uncoated, positive, negative and PEG coated. Also CNTs, Nanocellulose and more.	Omics and intrinsic properties on NM <i>in vitro</i> / <i>in vivo</i> effects <i>in vitro</i> high throughput screening	>100 NMs 31
SmartNanoTox	TiO ₂ , SiO ₂ , Au, carbon nanotubes etc. (binding free energies and potentials of mean force for interactions for all)	Interactions of amino acids and components of lipids and sugars with NMs (computational and experimental data)	>5 NMs
NanoFASE	TiO ₂ , CeO ₂ , Ag, Ag ₂ S	Transformations of NMs in the environment (air, water, sediment, soil, waste treatment and biota) and release models	
Nanomaterial Registry	Diverse NMs	NanoMaterials Registry Database	>2000 datapoints
NanoTEST	TiO ₂ , two fluorescent SiO ₂ , Iron oxide coated and uncoated, PLGA	Genotoxicity, cytotoxicity, uptake, oxidative stress	>5000 datapoints
S2NANO	Various engineered NMs (e.g., Oxide NM, Metallic NMs, and Carbonaceous NMs) 28–30 Curated from literature and experimental studies.	PChem properties / characterization; cytotoxicity assay conditions.	16 NMs datasets
CEINT NIKC	Ag/Ag ₂ S, CuO, Graphene oxide, CNTs, CeO ₂ , nZVI, cellulose nanocrystals, TiO ₂ , gold etc. – literature curated datasets / mesocosm datasets.	NM intrinsic, extrinsic (system dependent), social (e.g. use scenarios, matrix, concentration in products) properties; System characteristics; Exposure / Hazard data; Meta-data (protocol, temporal and spatial descriptors etc.)	20 NMs datasets
UC-CEIN NanoDatabank	Pristine MOx NM, quantum dots, CNTs/graphene 300 toxicological assessments, 150 investigations (curated data from over 500 publications)	PChem properties / characterization; toxicological assessments; NM fate, transport and material characterization.	>1000 NMs
Modern	metal oxide NPs of 12 sizes between 5 and 60 nm	35 full particle nano-descriptors	24 NMs 10,080 datapoint
NanoTOES	Ag NMs: 3 different sizes same surface properties, Ag NMs 20 nm with 6 different surface properties	PChem properties / characterization; cytotoxicity and genotoxicity endpoints;	9 NMs, >1000 datapoints
eNanoMapper	Publicly available datasets included	PChem properties / characterization; Hazard data	636 NMs, ~1750 datapoints

strengths/weaknesses of intrinsic and extrinsic properties (where they go, i.e. persistence, or, what they do, i.e. reactivity) in terms of predicting NMs behaviour [151]. The major intrinsic properties that they focused on included particle size distribution (number average), particle shape (e.g. aspect ratio), surface area, redox potential/band gap, crystalline phase(s), hydrophobicity, chemical composition (impurities, surface chemistry), and, rigidity. The extrinsic properties related to persistence included biodurability, zeta potential, density, dustiness (depends on moisture), dissolution rate (in environment, acellular), agglomeration/hydrodynamic diameter (dispersion stability) and surface affinity. Those related to reactivity only included reactive oxygen species (ROS) production and photoreactivity. A key factor in determining the extrinsic properties is the need to characterize the NMs in the relevant exposure medium and across the relevant exposure times [87,89].

2.1.2. Computational approaches for data gap filling and interpolation

Technological advances including high content and high throughput screening and omics approaches have transformed nanosafety research into a data rich field. Concurrently, nanoinformatics and ML-based *in silico* modelling applied to nanosafety require even larger data sets. For NM property models to be robust, predictive, and broadly applicable, large amounts of high-quality and complete experimental data are needed, that are organized and accessible. A current bottleneck is the fragmentation and inaccessibility of much of the data generated to date. To overcome this data fragmentation and facilitate model development, new processes need to be developed and implemented that will allow the capturing of both the NM data and the associated metadata making the produced datasets findable, accessible, interoperable and re-

usable (FAIR) [54,181]. This two-fold process aims at extending existing nanosafety and nanoinformatics databases with interfaces that allow curation, ontological annotation and semantic mapping of the data schema, and extraction of data and knowledge about NMs, as well as implementation of very focused set of experiments, designed to fill gaps identified in the existing large datasets. The combination of these activities will support the development of computational predictive toxicology methods, enable *in silico* models to perform at optimum levels, and increase the predictive power of the overall Integrated Approaches to Testing and Assessment (IATA) for human and environmental RA that is envisioned within NanoSolveIT. A strong focus is given to the production of curated, reliable NMs safety data under the FAIR principles [181].

Among the approaches available to overcome data fragmentation and data inaccessibility is data mining from literature and subsequent meta-analysis of the curated data [5,71], although this can be extremely time consuming if done manually. Text mining algorithms are under development [44,78] and their suitability for NMs is being evaluated within other ongoing nanosafety data projects such as [NanoCommons](#). One issue arising from these literature mining approaches is the degree of data quality and completeness [83,92], as well as the comparability of datasets generated by different groups using different batches of NMs [102] or characterization protocols. More specifically, data quality concerns originate from the different methods/protocols and experimental conditions used for the data production (e.g. lack of characterization in the exposure media) and the lack of sufficient metadata to describe the data and ensure interoperability. Data completeness concerns are linked with the amount of PChem characterization of NMs performed prior to and during the experimental procedure in the rel-

evant exposure medium, and the respective endpoints (toxicological, omics etc.) measured. The impact of these factors on modelling can be increased data variance and unreliability and decreased model robustness and predictivity. This is especially true in the development of nano-QSAR/nano-QSPR (quantitative structure–property relationship) models, where the variability observed in the PChem properties of NM might be higher than it is in reality [83].

To overcome these issues several approaches have been proposed [36,37,45,90,168]. For example, Ha et al. [45], demonstrated the benefits of data gap filling during the meta-analysis of the cytotoxicity of metal oxide NM using data mined from 216 publications, which resulted in 6,842 data rows and 14 attributes of nanostructure descriptors, PChem, toxicological and quantum–mechanical (QM) (computational) properties. Gap-filling was achieved using information from manufacturers' specifications, references utilizing the same NM or with estimation from other PChem properties. Data quality was assessed using a scoring system based on the presence and origin of data. Gajewicz et al. [36], on the other hand, proposed the use of read-across algorithms to predict the missing values and improve the predictive outcome of predictive models. In all cases, data gap-filling and increased data quality resulted in increased accuracy of the nanostructure–activity relationships models.

For such methods to be successful, detailed workflows for the experimental and literature data curation are needed [130]. Similarly, standardized experimental workflows, and complete reporting of experimental procedures including computational and database mining, need to be established to ensure data reproducibility at a wider scale [17,76,137]. NanoSolveIT aims to develop the gap-filling approaches and curation workflows through the gathering of existing datasets originating from recently completed EU funded projects (e.g. NanoMILE, NanoFASE, caLIBRAte, NanoTEST and others reported in Table 1 above), performing the necessary evaluation of the existing data and metadata and designing detailed experiments to fill the identified gaps in the NM PChem characterization and toxicity endpoints, thus increasing their quality. The resulting larger datasets will then be used by the modelling partners for the development of more robust *in silico* approaches with wider domains of applicability and enhanced predictive capability, while the developed standardized workflows will be made available for experimentalists to guide them in the production of the necessary scale and completeness of data for use in modelling approaches.

2.1.3. Dedicated NMs databases organized for modelling and informatics

The NanoSolveIT knowledge base will extend the NanoCommons / eNanoMapper databases with innovative, ontology-based application programming interfaces (APIs) to allow semi-automated curation and extraction of data and knowledge about NMs to support development of computational predictive toxicology methods. It will cover a wide range of data that researchers are looking for: omics data, nanodescriptors and relevant literature, as well as PChem properties and biological effects. There are two key aspects of the knowledge base. Firstly, NMs-specific datasets will be federated with other databases that are optimized for endpoints from proteomics, transcriptomics etc. for which well-established data management and deposition solutions already exist. The NanoSolveIT knowledge base will communicate via APIs and integrate the data via semantic mapping of their data schemas. Secondly, the knowledge base will be enriched by integration of data from protein structures, known signalling pathways, crystal structure information, for example. These data will be used by physics-based modelling approaches (see Section 2.3) to computationally design NMs and their biomolecule fingerprints. Other

types of data that can be enriched in the knowledge base include environmental data such as river pH, ionic strength, dissolved organic matter etc. which can support development of enhanced models for prediction of NM's environmental transformation and consequent ecotoxicity.

Furthermore, the NanoSolveIT knowledge base will adopt Open Science approaches to trigger open innovation with related projects and future users and collaborators. This will result in the NanoSolveIT Knowledge Infrastructure containing curated, reliable NMs safety data, accessible and reusable within the project, by the nanoinformatics community and by all stakeholders. The NanoSolveIT knowledge base will be helpful for those researchers interested, not just in the structural characteristics (descriptors) and PChem properties, but also in the known biological effects of particular NMs. Researchers can use the interface to navigate through the available data or retrieve data for further exploitation.

2.2. Toxicogenomics modelling (predictive models using omics data)

Toxicogenomics modelling is a subdiscipline of pharmacology that deals with information about gene and protein activity within a particular cell or tissue of an organism in response to exposure to toxic substances. It can be used to link the safety of NMs to the underlying biological mechanisms of their toxicity. Gene expression data can be combined with biological pathway information to identify possible adverse outcomes [43,67,110]. In order to make sense of the large volume of data generated by bioinformatics analyses, SOPs for the interpretation of the results in the correct biological context need to be established [46]. Adverse Outcome Pathways (AOPs), which describe in mechanistic detail the sequences of events that are necessary for an exposure at cellular and subcellular levels to lead to an adverse event or outcome at the organ or organism level, are a useful tool for organizing bioinformatics and other types of results into a predictive framework [111].

During the last decade, multiple efforts have aimed at characterizing the mechanism (mode) of action (MOA) of toxic chemical exposures using transcriptomics profiling of the exposed biological systems. This generated large reference data sets such as connectivity map [152], TG-GATEs [51], DrugBank [184] and LINCS L1000 [66]. These have been extensively used for drug repositioning (e.g. [53,106]) and toxicity prediction (e.g. [67]). The general concept of this approach is that toxicogenomic experiments identify the primary molecular changes in cells and tissues as a direct consequence of toxin exposure, and hence directly inform the toxicity pathways of tested compounds [43]. As an example, Predictive Toxicogenomics Space (PTGS) components, which were developed based on connectivity map data to predict organ toxicity, are likely useful as descriptors of AOP-linked MOAs and key events (KEs) in the affected signalling pathways [67].

Furthermore, when multiple time points are screened, toxicogenomics data can give a robust insight into the toxicokinetics and can further assist the drafting of an AOP [35,111,185]. Toxicokinetics describes the absorption, distribution, metabolism and storage/excretion of chemical toxicants, while, toxicodynamics describes the adverse (biological) effects that a toxicant has on an organism, e.g. altered structure/function and disease. Both processes are determined by the structure (morphology) and PChem properties of NMs, e.g. size, shape and surface reactivity [95]. Toxicogenomics data can also be exploited to infer similarities between different types of exposure and between exposure and human diseases [145]. Finally, new opportunities are emerging for integrating 'omics data with intrinsic NMs exposure properties to allow hybrid quantitative structure and MOA activity relationships to be developed [146].

More recently, toxicogenomics approaches have been employed to describe the MOA of NMs in various exposure scenarios *in vitro* (e.g. [63,101,144]) and *in vivo* (e.g. [47,52,64,70,107,132,141]). When multiple doses of toxin are screened by 'omics technologies, dose-dependent events can be extrapolated in order to further dissect specific mechanisms of toxicity [133,158]. In fact, dose metrics are a basic requirement for any *in vitro* screening to assess potential health risks of NMs. Genomic dose responses can be used to define the biological potency of a material as well as points-of-departure concentrations denoting adverse levels of exposure to the organism or cell. Typically point-of-departure concentrations are set using concentration response modelling or the lowest observable effect level approach [158]. Although transcriptomic response itself is often triggered as a response to an adverse reaction to the environment, a pathway-level concentration is typically used, as this value is more robust than a gene-level response and is directly connected to a known biological response. Further modelling methods, such as the PTGS, can be helpful to differentiate between adverse, adaptive and benign (or even beneficial) biological responses to chemicals. Mechanisms or pathways connected to key event responses in known AOPs are also applied for selecting adverse responses among bioinformatics analysis results [111]. When cell culture data is utilized, there is also the need to extrapolate the cell culture concentration to a biological exposure scenario which in the case of NMs is typically inhalation-based [111]. However, focusing on mechanistic aspects, the pathway-level benchmark concentrations can be used to rank NMs based on biological potency. Taking the concentration response into account is also helpful for biological grouping, e.g., for selecting optimal treatments for connectivity mapping-based biological similarity analyses. As the cost of toxicogenomics is steadily being reduced, its use in safety assessment and mechanistic analyses will only grow in importance.

NanoSolveIT is collecting existing toxicogenomics data, then transforming, analyzing and modelling it. The project has multiple aims, including deriving signatures of NM MOAs to support AOP

development; generating useful predictive models of the biological effects of NMs; and developing computational methods and software to be included into a robust computational platform for *in silico* nanosafety analysis (as seen in Fig. 1).

2.2.1. Deriving signatures of NM MOA to support AOP development

Toxicogenomics has provided unprecedented opportunities to clarify the MOA of many chemical exposures. The essential idea is to identify a set of genes that are significantly altered in a given biological system of interest due to exposure to a toxic agent. While conventionally, 'omics data analysis resulted in lists of differentially expressed genes, these are not *per se* informative of more complex patterns of regulation that underlie broader biological functions. To elucidate these functions, systems biology approaches, such as gene network reconstruction and inference, have been used to identify complex patterns of molecular regulation and co-regulation [64,192]. Moreover, the systematic annotation of molecular changes into known biological pathways has helped define toxicity and other AOPs [70,111,144]. To date, the transcriptome has been the primary molecular focus of such studies, followed by proteomics and metabolomics. Multi-omics approaches have already been used extensively to generate more thorough landscapes of molecular changes in human diseases (e.g. The Cancer Genome Atlas) and are beginning to be used to build more general models of NM MOA [144]; [143]. Omics-derived biosignatures are valuable for comparing the effects of different types of exposure, such as NMs and small molecules that would otherwise be difficult or impossible to detect by other means. For example, omics-derived NM biosignatures have been systematically compared to those from small molecules, drugs and human diseases in search of direct exposure-disease associations [145]. These analyses have identified biomarkers useful for biochemical assays in the zebrafish model, enabling the drafting of an AOP for metal and metal oxide NMs impacts on the central nervous system [75].

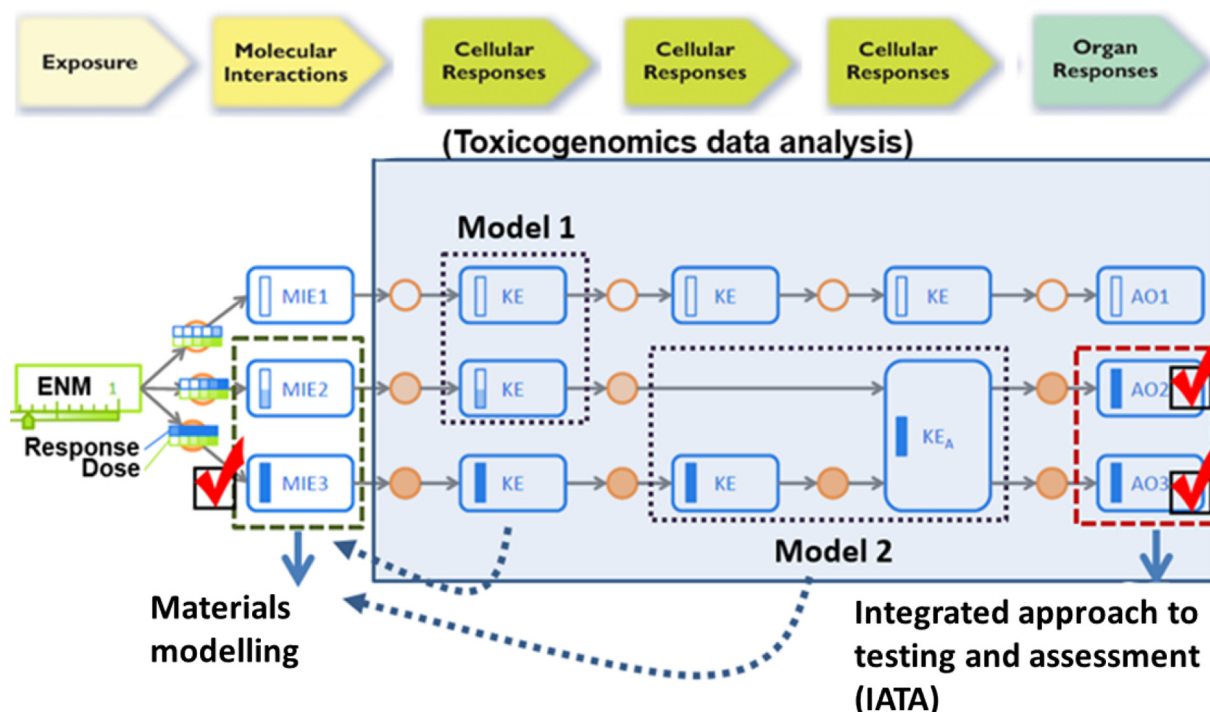


Fig. 1. Schematic overview of the workflow for toxicogenomics modelling and how these models feed into the subsequent materials modelling and IATA. AO – Adverse Outcome; ENM – Engineered Nanomaterial; KE – Key Event; MIE – Molecular Initiating Event.

2.2.2. Deriving signatures of NM MOA that define robust predictive models based on NM bioactivity

Toxicogenomics data has more recently been used to generate predictive models of toxicological and pharmacological interest. Multiple strategies have been employed: identification of specific biomarkers [145], or discovery of broader gene sets with strong predictive ability (PTGS, zebrafish-based toxicogenomic space, etc.). Typically, 'omics data analyses use univariate statistical testing, where each molecular feature is tested for significant differences between exposed and unexposed sets in replicate samples. However, these approaches allow derivation of only very limited individual or linear or concatenated biomarkers. The output from these analyses, importantly, does not guarantee biomarker specificity, although they are often referred to as such in the literature. Thus, more sophisticated feature selection strategies are needed, that allow non-linear combinations of molecules, or sets of biomarkers that are more specific. To this end, a number of algorithms have been proposed including GALGO [138] DIABLO [149], MANGA [145], MLREM [138,167,9,34,6,147].

2.2.3. Developing computational methods for inclusion into a robust computational platform

Despite the great value of toxicogenomics approaches in identifying important (NM) toxicity mechanisms in an unbiased manner, these approaches have thus far struggled to be implemented in the mainstream regulatory framework for chemical safety assessment. One reason for this is that toxicogenomics data are usually difficult to interpret without strong skills in bioinformatics and biostatistics. Importantly, the scientific community has been successful in developing improved analytical methods but has not yet agreed on formalized standard analytical SOPs. Clearly, omics data analysis is used extensively in other biomedical research fields, so many of these standard methods should be transferrable to NM toxicogenomics problems. The NM toxicogenomics community needs to ensure that this existing expertise is converted into standardized pipelines and software for nanosafety applications. Examples of useful methods are the eUTOPIA software for omics data preprocessing [94] and InFORM for gene network inference [93]. Similar efforts are being undertaken, within NanoSolveIT and several other EU projects, to further resolve dose dependent patterns of molecular change and to benchmark the resulting toxicogenomics and AOP models to increase their utility and acceptance by the community.

2.3. Multi-scale modelling framework for NMs property prediction

Adverse human health effects can be triggered and modulated by molecular-level interactions at the bionano interface, i.e. a nanoscale layer where biological entities meet foreign materials. These interactions are often non-specific and unintended. The currently poor understanding of the bionano interface means that RA for NMs and biomaterials broadly is largely based on empirical evidence and not on the mechanistic action of the adverse effects. In general, NM properties primary responsible for adverse effects are largely unknown, or are not the same as the PChem properties that can be routinely measured [89,136]. Understanding these interactions and the bionano-interface structure will assist with developing safety regulations and reducing the associated health risks but also with achieving improved control over the surface activity in nanotech-based applications.

Steinhauser and Sayre [151], reviewed the OECD guidance documents for NMs risk assessment, which provided recommendations regarding the measurement and assessment of occupational exposure, consumer exposure, environmental fate, ecological effects and biokinetics, as well as considerations for *in vitro* testing of NMs, for increased reliability and relevance. However, of partic-

ular interest for NanoSolveIT were the guidance documents for exposure modelling and QSAR modelling. While QSAR models usually employ two-dimensional (2D) structural information from molecules, they can employ three-dimensional (3D) information also, making them suitable for predictions of NMs properties and behaviour. The benefit of 2D is that it provides a good visualization of the structure, where one can easily identify the connectivity of atoms, the presence of specific functional groups and predict reactivity. However, 3D focuses on the molecular level and includes additional information related to bond distances or angles, as well as connectivity or binding to ligands in relation to surface topography and can account for extrinsic properties such as formation of a protein (biomolecule) corona, for example.

2.3.1. Computational NMs descriptors

In addition to direct correlations between the NM structure (expressed in term of nanostructure descriptors), properties and toxicity, as described above, interactions at the bionano interface can initiate AOPs via sequestering or unfolding of proteins central to molecular initiating events (MIEs) and KEs of the corresponding pathways [21,32,86,88]. Although they may not be completely independent of the basic features of the NM (as expressed either directly by nanostructure descriptors or by their intrinsic properties), a systematic evaluation of these types of properties that express protein affinity, protein unfolding and potential formation of cryptic epitopes that can induce new signalling pathways [86], may make predictive models more compact and robust.

Simply stated, a *descriptor* is the final result of a logic/mathematical process that transforms chemical-based information (encoded within a symbolic representation of a chemical structure) and, in the case of NMs, also physical-based information (i.e. morphology) into a useful number that can be exploited by a predictive model. Thus, descriptors provide unique information required to draw (or build a molecular model of) a NM. Whereas *property* (e.g. solubility) should be considered as a consequence of the NM's structure; it is impossible to deduce back the structure from such properties only. The properties can be either measured experimentally or computed with use of first-principles-based methods (e.g. *ab initio*, Density Functional Theory, molecular dynamics). The correct distinction between *descriptors* and *properties* is important, since only the structure (descriptors) can be directly controlled by a designer in the safe-by-design process.

There are various levels for chemical structure representations (descriptors) ranging from one-dimensional (1D) descriptors (e.g. basic molecular formulas), to 2D descriptors (e.g. connectivity index), as well as 3D that are conformation dependent (e.g. dihedral angles, radius, shape).

Examples of computational properties based on NMs interactions that can be related to MIE, KE or AOPs include: composition of the NM protein corona; adsorption enthalpy of amino acids, lipid molecules, or proteins onto the NM surface; adsorbed protein or NM hydrophobicity; production of ROS, dissolution of NMs leading to release of ions, all of which must be determined in realistic environments [20]. Calculation of properties based on a full-particle molecular model, using *ab initio* quantum chemical (QC) or even semi-empirical methods remains unfeasible in the near future due to the large size of NMs and thus the associated enormous computational time needed. Therefore, a significant effort was invested by the community to develop different approximations and simplified molecular models of NMs to derive NMs properties.

One of the first approaches for calculation of nanodescriptors was the design of optimal molecular descriptors by Toropov et al. [162,164,166]. These descriptors are calculated from SMILES structures and consider the chemical composition of NMs. Information about experimental conditions of NMs synthesis can be included

in the descriptor calculation. This type of descriptor has been successfully used to model the toxicity of NMs [162,166].

QC calculations based on small clusters of atoms can be used to obtain such properties as HOMO-LUMO gap (band-gap between conductance and valence electrons) and enthalpy of formation, which is currently not possible for full-sized NMs. QC properties can be directly used to model toxicities of NMs without taking into account the size dependency of the properties [131] or can be extrapolated to obtain properties values for specific size of NMs [56]. To take into account the effect of size on the properties of NMs, the calculations based on so-called full-particle molecular model should be performed. Such type of calculations for metal oxide NMs have been performed by [157,155,91,11]. Full-particle properties were derived from molecular mechanics calculation of NMs and describe the energetics (potential energies), coordination numbers and other attributes of the NMs. Full-particle descriptors and properties have been successfully used to model the toxicity of metal oxide NMs [156].

Surface modifications, such as change in coating materials, influence the properties of NMs and should also be included in the modelling process. Xia et al. [186] proposed a biological surface adsorption index to describe competitive adsorption of proteins onto the surface of NMs [187]. The adsorption coefficient is expressed as a logarithmic function of five descriptors: excess molar refraction (representing molecular force of lone-pair electrons); the polarity/polarizability parameter; hydrogen-bond acidity and basicity; and the McGowan characteristic volume describing hydrophobic interactions [186]. Experimentally obtained log K values can be used to derive five descriptors for surface forces related to adsorption. Combining all of these approaches would allow the characterization and modelling of NMs in biological systems accounting for all important aspects (electronic effects, size dependency and surface modifications) and would greatly improve the quality of nanoQSAR models.

2.3.2. Modelling (predicting) NM corona composition

The structure of the bionano interface can be simulated using first principles physics-based methods. Such simulations are often computationally intractable or use model systems that do not capture the complexity of real biological environments [72]. The relevant system sizes are too large for direct atomistic simulation, so the properties of interest can only be accessed using a coarse-grained (CG) representation, where sub-nanometer interactions have been integrated out. Molecular details of the NM are preserved when the CG model is constructed using a multistep approach, where each layer is parameterized from simulations at finer resolution [129]. Atomistic simulations also have challenges as they rely on accurate and validated force fields. Where such force fields are available, all-atom Molecular Dynamics (MD) methods can be used to construct a united atom representation of the NM and associated biomolecules. Computational study of new materials requires development or optimization of new atomistic force fields based on QC calculations or experimental data [8].

Coarser scale simulations also have challenges due to the nature of biological samples: the number of relevant biomolecules interacting with the NM can be enormous; for example, human plasma contains over 3,700 proteins and even larger numbers of metabolites [16]. The corona composition (lists of proteins and metabolites (lipids and other small molecules) known to interact with a specific NM) may therefore be an impractical property to be used for predictions, although meta-analysis of over 63 NMs plasma corona studies suggested that about 125 proteins form the interactome of NMs [176]. Each NM immersed in plasma typically has its own unique corona that may involve hundreds of different proteins [13,23]. Proteins in the corona reflect the functionalities on the NM that bind specific types of biomolecule [85]. This changes

over time, as the most abundant proteins bind first, and are subsequently replaced by less abundant but more tightly bound proteins, and the corona also evolves as the NMs are internalized into cells for example [84], and the cells respond to the presence of the NMs [2]. Capturing this complexity in descriptors used in ML models is very challenging, and often statistical properties of descriptors for the proteins are used.

Early examples of corona-based predictive schemes exist in the literature. An extensive gold NMs protein corona dataset generated and analyzed by Walkey et al. [177], was re-analyzed in order to identify and quantify the relationships between NM-cell association and protein corona fingerprints in addition to NM PChem properties [1,80]. QSAR models were developed based on both linear and non-linear support vector regression models making use of a sequential forward selection of predictors. For example, an initial pool of 148 predictors was used, with the analyses eventually identifying 10 corona proteins and 3 PChem characteristics (NM size and zeta potential in cell culture medium) as the most significant factors correlating with NM cell association [1]. As more data emerges, including on the small molecule or metabolite corona and how these interact with proteins to form the complete corona [16], refined models and more detailed predictions of the composition and impact of the NM corona will emerge.

NanoSolveIT recently proposed a multiscale modelling scheme that enables modelling of large molecular assemblies in both length and time domains, which is not achievable by traditional atomistic simulations. This allows information on NM and biomolecule specificity to be preserved [82,81,129] whilst also enabling calculations in reasonable times. The NanoSolveIT method (shown schematically in Fig. 2) uses a systematic CG method that includes:

- parameterization of the atomistic force-field for the NM;
- calculation of interactions of the biomolecule building blocks (amino acids, lipid segments, DNA bases) with the surface of the NM and interaction between the building blocks at the atomistic level under specified conditions;
- parameterization of the CG force field for biomolecule building blocks and construction of a NM of arbitrary size and shape;
- CG modelling of interaction of entire biomolecules with the NMs' surface and calculation of preferred orientation and the mean adsorption energies for the bound biomolecules;
- further coarse graining for lipids and proteins to make united amino acid blocks and study competitive adsorption and bionano-interface structure.

The multiscale modelling framework proposed by NanoSolveIT enables calculation of descriptors and properties of the bionano interface for a large number of biomolecules in a short time. The new properties include those for proteins (principal moments of inertia, charge, dipole moment, hydrophobicity indices, and the solvent accessible area) and those for interaction with NM (Hamaker constants for residues, their mean adsorption energies, and the overall adsorption energy for the protein globule or lipid molecule) [82]. These properties will be used to produce interaction fingerprints for arbitrary NMs with respect to specified biological activities and will thus provide key information for toxicologically relevant predictive modelling, e.g., for predicting NM ability to induce an AOP via MIE or KE.

Several models describing how binding to NM surfaces affects protein conformations and subsequent recognition behaviour have appeared in the literature. These will be very useful for connecting corona composition to molecular initiating and other key events in AOPs. Multiscale MD simulations of a single NM with the protein ubiquitin demonstrated that ubiquitin competed with citrates for the NN surface. At a high protein/NM stoichiometry, ubiquitins formed a multi-layer corona on the particle surface, with the pro-

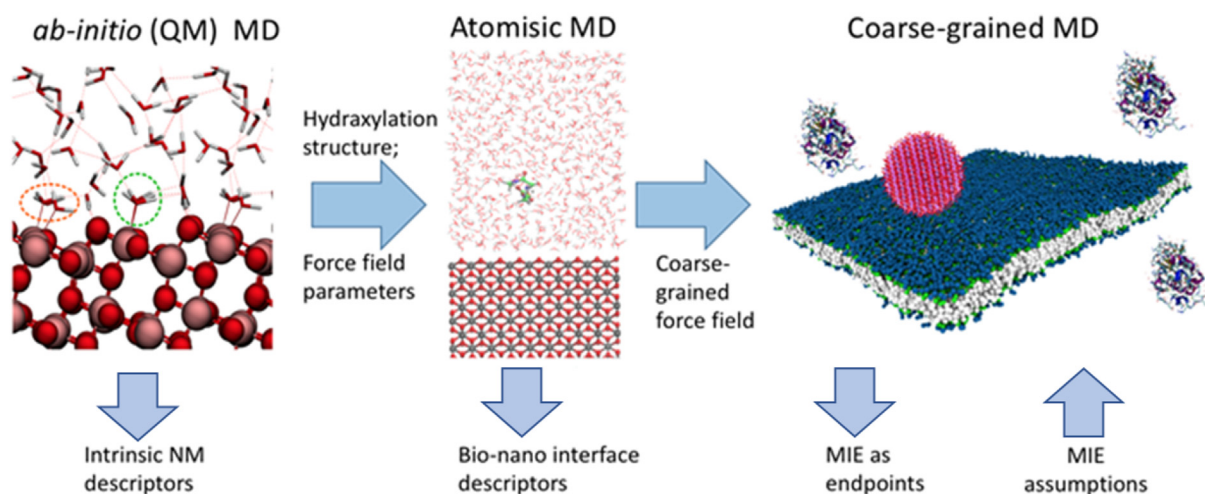


Fig. 2. Schematic illustration of the NanoSolveIT approach to multi-scale modelling of NM interactions with biomolecules to form the biomolecule corona which provides the biological identify to the NM and determines its subsequent uptake and impacts in cells and organisms.

teins exhibiting conformational changes that included destabilization of α -helices and increased β -sheet content of the proteins [22]. A significant challenge with this approach is that it is unknown whether this protein binds under competitive conditions (e.g. in plasma), and whether these conformational changes occur in real systems. However, the correlation between unfolding of a specific protein and receptor activation has also been demonstrated experimentally and modelled using MD. Ding et al. showed that specific sizes of negatively charged poly(acrylic acid)-conjugated gold NPs bound to, and induced unfolding of, fibrinogen (Fg). This promoted interaction with the integrin receptor, Mac-1, leading to increased NF- κ B signalling and release of inflammatory cytokines [21]. Building on this work, Kharazian et al. used MD simulation to investigate how poly(acrylic acid) coats a gold NM surface. The root-mean-square deviation (RMSD), radius of gyration (Rg) and solvent accessible surface area (SASA) properties from the calculations showed that the gold surface can induce Fg conformational changes favouring an inflammation response [62]. They suggest that the integrity of coatings on ultra-small gold NMs are compromised by the large surface curvature, and that surface coatings may be degraded by physiological activity. Other modelling studies have also assessed biomolecule conformation in NM coronas and, where relevant, these approaches will be integrated into the NanoSolveIT toolbox.

2.3.3. Connecting NM-coronas to biological impacts/AOPs

Data exchange and reuse puts significant limits on how we represent the biology and chemistry of NMs. For example, to compare the risk of two NMs, it is necessary to know whether they are chemically similar and if they are behaving biologically in the same way (e.g. via read across studies). To ensure data reuse is possible, the NM data representations must be interoperable, i.e. both humans and machines must be able to make such chemical and biological comparisons. The bioinformatics and cheminformatics communities have developed extensive methods to perform such tasks. To be reusable, data must meet community standards around data quality [92], be interoperable (i.e. be able to interact with other databases), be machine accessible (i.e. be annotated to allow computers to understand the individual datapoints), not be hidden behind firewalls and be findable by search engines and models. Recently, these ideas were summarized in the FAIR principles [181] and the need for interoperability and data linking was recently outlined in a position paper [60]. A key step for FAIRness

was the development of a common NMs ontology to facilitate data interoperability [48,100].

Several recent studies focused on developing models of NM-related properties and biological effects [10,123,142,159,160,161,163]. For example, a set of 18 NMs were studied by Wang et al. [178] and showed that factors such as metal content, surface charge, and particle morphology induce high toxicity. Melagraki et al. developed, a predictive classification model based on OECD principles, for the toxicological assessment of iron oxide NMs with different cores, coatings and surface modifications based on a number of different properties including size, relaxivities, zeta potential and type of coating [97]. More recently a predictive nanoinformatics model, validated according to the OECD principles, has been developed for the prediction of the protein binding and the cytotoxicity of functionalized multi-walled carbon nanotubes [172].

Numerous other as yet-unexplored features may also be important for generating adverse effects from NMs. For instance, there is evidence to suggest that NMs, especially in the lower nm range, penetrate biological membranes and are able to reach organs that are otherwise inaccessible for larger substances [14]. Exposure to specific NMs has also been demonstrated to cause adverse biological effects by increased production of ROS, such as oxyradicals [79,103,112]. However, some NMs may be less harmful than their corresponding bulk forms in some instances [61] and even two NMs from the same source, with similar sizes and chemical compositions may exhibit diverse effects [121]. Clearly, factors other than size, shape and surface area are also important in controlling the interactions and effects of NMs in biological systems. For example, Zhang et al. proposed that the higher toxicity of fumed silica relative to Stöber silica stems from the formers' intrinsic population of strained three-membered rings along with its chainlike aggregation and hydroxyl content [189].

Surface coatings are crucial for control of both useful and adverse effects of NMs. They provide a great opportunity to improve materials through rational design, by enhancing a useful property and reducing adverse effects. This approach has been demonstrated for multi-walled carbon nanotubes (MWCNT) and asbestos fibers, where structural similarities suggested potentially harmful effects according to the so-called 'fiber paradigm'. The similarities in the two structures guided a pilot study, which showed that when long MWCNTs are injected into the abdominal cavity of mice, asbestos-like pathogenic effects can be induced [125]. However, less rigid forms of CNTs had lower toxicity [105],

suggesting rigidity as a potential design parameter. A larger pulmonary toxicological study of 10 MWCNT demonstrated increased inflammation and lower genotoxicity with increased surface area (decreasing CNT diameter) and some reduction in inflammation with –OH and –COOH surface functionalization [126]. Toxicogenomic markers of MWCNT-induced fibrosis [128] and acute phase response as a predictor of cardiovascular disease [127] have also been identified. Thus, future experimental and modelling studies will explore the correlations between a wider range of NMs parameters, functionalizations, corona compositions, cellular associations and AOPs.

2.4. Predictive nanoinformatics modelling (using artificial intelligence methodologies)

Increasing use of NMs has escalated concern about their potential risk to the environment and human health. The RA of NMs has traditionally been based on a variety of *in vitro* and *in vivo* assays, which are typically expensive, labour-intensive and time-consuming. *In vivo* animal testing has been constrained by ethical considerations to follow the 3Rs rule aiming to replace, reduce, and refine the use of animals for scientific purposes [140], and from 2013, the use of animals is banned for safety assessment of cosmetic products [28]. Due to the multitude and variety of NMs increasingly exploited in our daily life, it is impossible to assess *in vitro* or *in vivo* potential risk of all NMs on a case-by-case basis [33]. In addition, application of the full REACH information requirements for every single variant of a given NM regarding particle size, shape, or surface properties, which accelerate the toxic nature, would lead to an insurmountable amount of testing [30].

To complement existing toxicity tests, computational approaches (*in silico* models) have been proposed as useful alternatives to predict *in silico* the potential hazard of NMs, to reduce time and cost of nanosafety assessments and provide input for design of safe and functional NMs. QSARs, as *in silico* models, have been developed to predict biological activities, including the toxicity of various substances [134], mainly based on molecular structure [96], SMILES [179], and international chemical identifiers (InChI) [49]. However, such classic QSAR approaches have shown many limitations in predicting the toxicity of NMs due to the lack of standardized databases on their structures together with PChem parameters/descriptors and various toxicity endpoints [119]. In addition, as the toxic effects of NMs vary according to their PChem properties and the same type of NMs exhibit diverse toxic effects under different biochemical conditions (e.g., cell line, cell species, etc.) [39,124,191], making classic QSAR modelling difficult.

To overcome these challenges and establish relationships between nano-descriptors, which express the novel and size-dependent properties of NMs, and toxicological adverse effects triggered by NMs, various modelling approaches based on statistics and machine learning (ML) have been proposed to predict qualitatively or quantitatively *in silico* endpoints. Generally, developing predictive models includes the following major steps [114]:

- (1) data collection that contains associations between NMs and endpoints;
- (2) data preprocessing to transform the raw data into a useful and efficient format and to improve the quality of the raw data and reduce batch effects between experiments;
- (3) selection of nano-descriptors (fingerprint) predictively linked to functionality and hazard of NMs;
- (4) predictive model development and validation;
- (5) mechanism interpretation;
- (6) and definition of the applicability domain of the model.

Early ML models for predicting NMs properties, such as protein corona (PC), have begun to appear in the literature. For example, a random forest classification approach has been shown to model PC populations for an array of NMs properties and reaction conditions, while providing insight into feature importance to define which aspects of protein, NM, and solvent chemistry are the most important to defining the PC population [31]. The model has the potential for prediction of NM PC fingerprints across a wide range of NMs, protein populations, and reaction conditions. Varsou et al. presented a novel read-across ML methodology based on a mathematical optimization approach for the prediction of NMs toxic effects [170]. This is an area of intensive research currently and numerous advances are expected in the near future.

2.4.1. Calculated toxicity predictors for ML models

Effective prediction of NMs health and environmental risks requires utilization of existing or newly generated data to develop *in silico* hazard assessment models based on refined hazard-correlated endpoints. Achieving this requires establishment of a unified methodology for predicting the risks related to use of NMs, building on a sustainable multi-scale nanoinformatics framework, which links existing and emerging data and integrates, facilitates and advances the current state-of-the-art *in silico* modelling and predictive toxicology approaches.

A major objective of nanoinformatics is thus the development of *in silico* methods for predicting biological activities of NMs. These models will be trained on NM structural information (encoded by nanostructure descriptors) and PChem features (properties). Therefore, one of the most important and difficult challenges is to devise nano-specific descriptors that encode these features. When dealing with large numbers of NMs, it is important to use calculated descriptors, as those that require experimental measurement will become less useful and more expensive to generate as the data set sizes grow. Initial strategies for nano-specific descriptors include those derived from SMILES strings, the periodic table, simplex representations of molecular structure, liquid drop model descriptors and those derived from NanoJava applets [27,58,69,150,155]). Additional potentially valuable information about morphological features of NMs will be extracted from microscopy images, e.g. size, area, circularity, and aspect ratio, surface topography [113] and the use of modern automatic image analysis software tools such as NanoImage and NanoXtract developed specifically for NMs. For instance, utilizing NanoXtract [171], a NM Image Analysis tool powered by the Enalos Cloud Platform, users can analyze TEM microscopy images and extract useful descriptors that can be used in a future step as inputs to predictive models. Within a simple and user-friendly interface, the user can upload a single TEM image of a specific NM and with just a few clicks to obtain a set of NM image descriptors displayed on the screen or downloaded as a .csv file. These image descriptors can be explored by developing nanoQSAR models, either within the Enalos Cloud platform or using in-house models, to identify those descriptors most predictive of NM behaviour and /or biological effects.

2.4.2. Meta models to integrate across scales

An optimum set of nano-specific descriptors will enable generation of models describing the relationships between them and experimental and computed NM properties (intrinsic and extrinsic) and a wide variety of biological endpoints. In classic nano-QSAR modelling, both structural information and properties are treated as independent variables. However, NanoSolveIT will also use meta-models, where a property (e.g. hydrophobicity, NM-

biomolecule interactions, surface activity, adsorption properties) is first derived from descriptors, e.g.:

$$\text{property} = f(\text{descriptors of NMs})$$

This meta model ML approach will also allow rapid estimation of properties of many NMs that would normally be obtained from computationally demanding QC calculations and/or molecular dynamics simulations. Furthermore, meta models will be used to provide deeper insight into the toxic effect induced by NMs, from the mapping between NM properties (obtained from simulation), uptake, internalization and Physiologically Based Pharmacokinetics (PBPK) considerations, and the observed toxic effects, as shown in Fig. 3. This will be achieved by analyzing which structural, PChem, or meta model-derived property has the largest impact on adverse biological outcomes induced by NMs. This approach also allows implementation of the safe-by-design concept, where ML models of commercially valuable properties and adverse biological effects are used to rationally design NMs with optimum performance and greatest safety. These methods are also valuable for research groups and institutions that focus more on NM synthesis rather than RA.

Thus, NanoSolveIT proposes a new, broader view of nanoinformatics. All ML models developed in the project will be trained on heterogeneous data from multiple sources (experimental characterization, image analysis, computational simulation, omics data) that have, until now, been treated individually. These integrated NM “fingerprints” will contain maximal information required to build statistically robust and optimally predictive models of biological activity. Optimal predictivity will be ensured by the use of sparse feature selection (MLREM, LASSO), Bayesian regularization to control model complexity (LTU, Biomodeller), and automated processes included in the RRegrs R package [169]. The RRegrs package automatically selects the next algorithm for each case from 10 different statistical and ML methodologies (e.g. multiple regression, lasso, support vector machines, and artificial neural networks). Computational efficiency will be achieved by the use of Graphical Processing Unit (GPU) algorithms through the *Enalos +* toolkit [173]. This improves big data handling and accelerates computation. All models will be built according to OECD guidelines, where feasible. All models will be validated using test sets, cross-validation, bootstrapping and Y-scrambling as appropriate [3]. Domains of applicability for models will be calculated by established methods, such as the leverage method and the Euclidean distance to the training set.

2.5. NM human and environmental RA

The ultimate goal of multi-scale nanoinformatics modelling is the generation of a computational platform that integrates *in silico* approaches to NMs safety testing and assessment into an overall risk framework or IATA. The IATA will provide a systematic framework for the integration of the data and models used and/or produced during the project's lifetime. The IATA will be configured to ensure correct mapping of the sequence with which the developed and integrated tools are run thereby ensuring that the inputs and outputs from the various tools are harmonized and interoperable. This will ensure proper tool function and optimization of the initial data input based on the desired outcomes in order to answer specific stakeholder needs (safe-by-design, exposure, hazard and RA) and make predictions regarding the safety, mode of action, and likelihood of a NM triggering a specific AOP. NanoSolveIT is building on previous efforts to map and evaluate a range of NMs-specific risk assessment tools or so-called “new approach methodologies” in terms of their ability to provide “added value” and “decision support” as well as their status in terms of whether they are ready for implementation or require further exploration and development [109].

One of the key functions of the IATA will be the development of NM fingerprints - predictive and informative PChem, biological and computational descriptors that describe a set functionality, i.e. the minimum set of descriptors required as input to predict specific NM functionalities. The IATA will focus on those parameters that are easy to measure/calculate on a regular basis during every day experimental and computational practice. Classification approaches of relevant descriptors as intrinsic (do not change irrespective of the exposure route, release pathway) or extrinsic (context-dependent) and the stability of potential transformations will be also incorporated into the IATA [89]. This will reduce the experimental/computational workload to derive specific descriptors and subsequently the amount of information needed to create robust predictions, especially in terms of safe-by-design. In the case where a NM is classified as hazardous, the generation of additional parameters will be proposed to refine the ultimate hazard and RA. As a result, a certain set of exposure and toxicokinetics models will be integrated and linked within the IATA. These models will act in a synergistic way to fulfil the requirements of the NanoSolveIT framework.

In every case, NanoSolveIT will look to deliver informative and validated predictive models with clear definitions of their domains of applicability. For the NanoSolveIT IATA to be successful, these

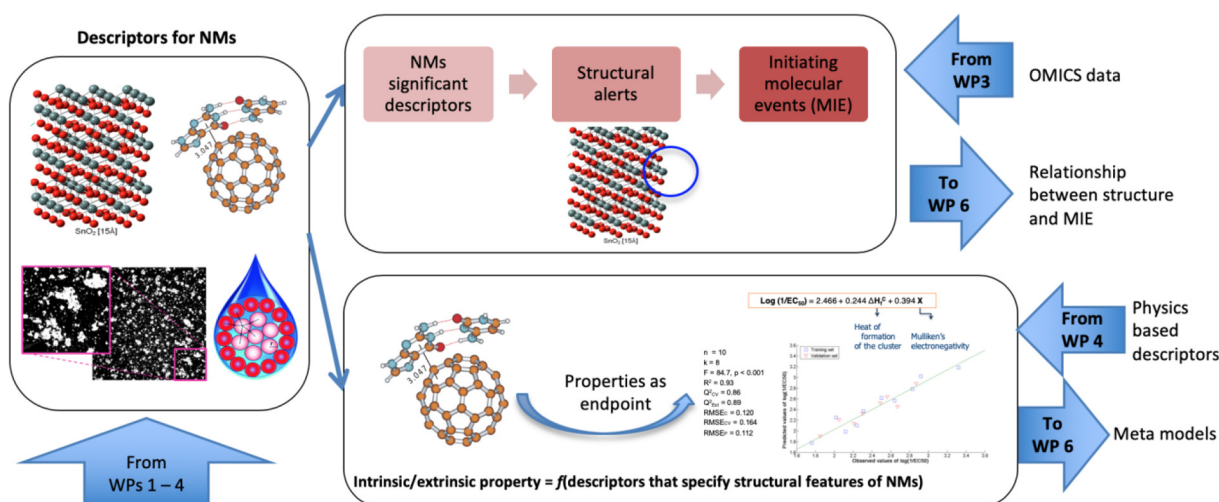


Fig. 3. NanoSolveIT meta models.

models need to act synergistically. This will require standardized modelling approaches and workflows that take into account the OECD validation principles [114] and recommendations from the European Materials Modelling Council [26].

To achieve the required synergism, a clearly defined workflow has been established, which includes the following steps:

1. Selection of modelling tools for the NanoSolveIT IATA, appropriate to the needs of stakeholders.
2. Evaluation of model parameterization needs to identify the required parameters for each tool to support the NM fingerprint and the IATA.
3. Generation of plausible and useful exposure and environmental release scenarios covering the entire lifecycle assessment of NMs. These will be based on data from recent, ongoing and future studies covering NMs' release points during production, use and disposal.
4. Technical solutions for incorporation of the various modelling tools into the NanoSolveIT IATA and e-Platform.

To achieve these ambitious goals and demonstrate the utility of the IATA to the research, industrial and regulatory communities, NanoSolveIT is building on recent advances from e-infrastructure projects OpenRiskNet and NanoCommons. For example, a key challenge is integration and alignment of different sources of information to allow use of weight of evidence approaches and evaluation of uncertainties in the information. Here approaches already implemented in NanoCommons will be leveraged. For example, NanoCommons has already integrated the similarity scoring and data quality modules of the GUIDEnano platform for RA and risk management of NMs, which are essential for weight of evidence analysis. Similarly, NanoCommons has supported the OECD WPMN project on NanoAOPs [117], including updating and expanding the database for the KEs identified for the tissue injury adverse outcome. It is building on that experience to develop KE-specific data capture templates for electronic notebooks, harmonizing data capture and data curation approaches, and text mining tools to speed up compilation of literature datasets. Integration of tools for searching the NanoWiki and identifying KEs is also currently underway in NanoCommons, and will be incorporated into NanoSolveIT. NanoCommons has already demonstrated approaches for integration of datasets, their curation and quality assurance, using its two main modelling platforms, namely Enalos tools via KNIME and Jaqpot via Jupyter notebooks. Currently these NanoCommons deliverables are under review by the European Commission but will be made publicly available via the [NanoCommons website](#) as soon as possible, to allow wider community adoption of the approaches, as per all the public deliverables. Datasets can be pulled from a range of existing databases such as PubChem, UniChem, UniProt etc. via dedicated Enalos APIs (through KNIME nodes) and Python and R-scripts for Jaqpot.

A final demonstration of the utility of the NanoSolveIT IATA will be via case studies that showcase the various components and the overall approach. These will include existing OECD IATA case studies and assessment of their applicability to NMs, as well as a case study on developing NanoAOPs for genotoxicity and potentially for accelerated ageing in *Daphnia magna* based on the existing extensive datasets available within the NanoSolveIT consortium.

The three OECD IATA case studies identified for NanoSolveIT are: –

1. Prioritization of chemicals using IATA-based Ecological Risk Classification (Prioritization of chemicals / Ecotoxicity) – NanoSolveIT will adapt for NMs;
2. Case study on grouping & read-across for NMs genotoxicity of nano-TiO₂ (Grouping / Genotoxicity) – NanoSolveIT will assess the

ability of our models / IATA to reproduce (improve on) the findings;

3. A Case Study on Use of IATA for Sub-Chronic Repeated-Dose Toxicity of Simple Aryl Alcohol Alkyl Carboxylic Esters: Read-Across (Grouping (Read-across) / Repeated dose toxicity) – NanoSolveIT will adapt this for NMs.

It is also worth noting that expert judgement is an essential part of the IATA process. Both main computational platforms being used to underpin the NanoSolveIT IATA, namely Enalos tools via KNIME and Jaqpot via Jupyter notebooks, are fully configurable as nodes. This allows the establishment of pre-configured workflows that need minimal input from the user, but that allow expert intervention and decisions where required. For example, workflows can include opportunities to download the raw and transformed datasets for quality assurance purposes, analysis of data quality reports, and other checks prior to proceeding to the next step. Additionally, Bayesian belief network are being integrated into the IATA to automatically manage expert judgement and modelling reliability. This allows analysis of the accuracy of the expertise and ensures that the expert opinion does not bias model expectations [19,148]. Bayesian networks are excellent quantitative and tractable RA tools that can handle and augment datasets with missing values by incorporating expert knowledge and judgement. They have been used successfully for developing IATAs for simple chemicals, and recently applications to NM RA have been reported [104].

2.6. NM cloud platform

A major outcome of the NanoSolveIT project is the e-platform that will be made available to the community as a cloud application accessible via the web or which can be installed as a standalone platform on local servers of interested stakeholders (see Fig. 4). This platform provides all the computational modelling tools and functionalities developed during the project under a common framework and with an optimized workflow for input of data and generation of predictions in response to specific stakeholder queries. These individual models and tools will also be available as separate components for specific applications and needs, but the ultimate goal is to integrate the tools as much as possible, to allow the implementation of complete computational IATA workflows.

The system architecture of the e-platform will be efficient, user-friendly, extensible, easy to maintain and open to the community for integration with other tools and databases. Our vision is to develop a central and easily extensible web-based platform of secured sustainability that will be customizable to specific user requirements and will address both current and future research and regulatory questions and needs.

To achieve all these objectives, state-of-the art tools and best practices in developing web applications and platforms will be leveraged. An open microservice architecture relying on the Open-API specifications for constructing and testing the APIs, and the REST (Representational State Transfer) architectural style for designing distributed systems will be utilized. This will consider relevant operational aspects of the platform such as extensibility, scalability, and elasticity.

The development of the NanoSolveIT e-platform will build on, incorporate and extend two of the most comprehensive platforms available in the nanosafety community, namely [Jaqpot](#) [18] and [Enalos](#) (NovaMechanics [109,171]). These are already used for developing, hosting, sharing and exchanging predictive models. Additionally, it will leverage exposure and fate models developed in recent H2020 projects, such as the NANOFASE model for environmental fate and effects, a multi-box aerosol-dynamic model

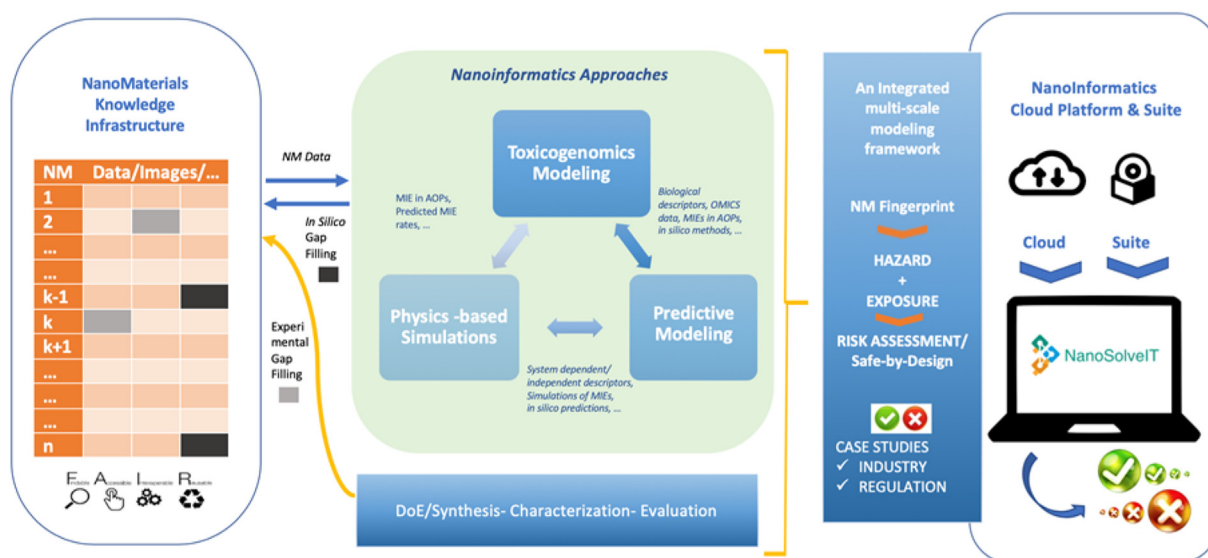


Fig. 4. NanoSolveIT platform integrating all of the various omics, materials and machine learning and meta models into an harmonized platform for NMs properties, exposure, hazard and risk prediction, safe-by-design and *in silico* NMs toxicology. All available data from the contributing projects and literature, including acute and chronic toxicity data, and a strong focus on regulatory-relevant endpoints, will be incorporated into the database.

[57] that with further development and testing will be used to calculate the size-resolved NM inhalation exposure and dose, and the GUIDEnano tool for exposure, hazard and RA which has already been integrated into the NanoCommons research infrastructure.

For simulation of the biokinetics of NMs, i.e. the administration, distribution, metabolism and excretion of NMs when they enter an organism, the NanoSolveIT platform will integrate and extend functionalities that have been developed in the Jaqpot modelling environment. Detailed PBPK / Toxicokinetics models will be provided as web applications, ready to simulate various NM external exposure scenarios (e.g. single dose, multiple doses, repeated dose etc.) and the physiological parameters of the individual or the group of people who are exposed. The concentration–time responses in the various organs and tissues of the organism will be automatically generated as interpretable and visual tables and graphs. By connecting the exposure and the biokinetics models, NanoSolveIT will develop integrated workflows that model how an external exposure scenario affects the internal exposure, i.e. the NM concentration in specific tissues of interest. By further integrating and interlinking these tools with models that can predict hazard at the AOP level, the NanoSolveIT platform will be able to predict whether an exposure scenario can initiate a specific AOP in a specific individual.

Hazard assessment services will be part of the NanoSolveIT platform, by providing ready-to-use web implementations of nano QSAR models and the read-across approaches that are being developed in the project. By integrating ML libraries, practically all state-of-the-art statistical and ML methodologies will be available for model building. Custom-made solutions, such as implementation of read-across methods or specific ML algorithms will also be possible if they are compatible with the architecture and the design principles of the platform. Hazards and RA tools developed in previous projects (e.g., NanoCommons & NanoMILE tools hosted on the Enalos Cloud Platform) are already available as user-friendly applications and will be further expanded. Several of the above mentioned predictive nanoinformatics models [1,97,98,172] are already hosted in the Enalos Cloud Platform and are publicly available. The Enalos Cloud Platform [174], is based on web semantic cheminformatics and nanoinformatics application, with the ability to host any predictive model as a web service with a user friendly user interface (NovaMechanics [108]). Web applications are

invaluable for computer-aided design of NMs since they can be used to predict the activity of new NMs prior to synthesis and the biological evaluation. Furthermore, ‘safe-by-design’ paradigms can be easily constructed for a systematic study of the effects of NM structural modifications and properties on several relevant biological responses and crucial properties.

Deep Neural Networks and deep learning algorithms have received much attention in the ML community and increasingly many other scientific, technological, business, and medical spheres over the last few years. These methods are closely related to or incorporated into big data analysis, especially for large volumes of high-resolution images. Applications of deep learning in the nanosafety discipline are so far quite rare. The NanoSolveIT project will fill this technology gap by developing infrastructure for implementation of deep learning models as web services. Specific demonstrator applications will be provided to showcase the capabilities of these technologies, for example use of only image data to predict whether an organism is affected or damaged by the presence of an NM. These AI capabilities of the platform will be supported by graphics processing units (GPUs).

The NanoSolveIT e-platform will be seamlessly integrated with the project knowledge base (Section 2.1.3) that will provide the necessary information for developing, testing and finally using the models for predictive purposes. All modelling components and the complete platform will comply with the ontological annotations used throughout the project and particularly in the development of the knowledge base. Model predictions will be used to fill gaps in the NanoSolveIT knowledge base, due to lack of experimental data. Thus, the system is designed in a manner that allows continuous improvement as new data emerges, and integrates seamlessly experimental and computational data, thereby constantly enriching the underpinning knowledge base.

2.7. Knowledge transfer and communication with stakeholders and RA bodies

Due to the current lack of *in vitro* and *in vivo* data risk assessors seek reliable *in silico* approaches, derived from validated or scientifically-valid *in silico* models, grouping and read-across, PBPK and toxicokinetic modelling [7], and for support scientifically

based regulatory nano-risk governance. The current state-of-the-art on alternative testing strategies in RA of engineered NMs was published by the OECD [116]. A tiered approach based on non-testing and *in vitro* methods has been suggested for the prediction of realistic biological outcomes when used in a weight of evidence (WoE) approach. Several frameworks and *in silico* approaches been proposed but even so risk assessors still consider *in silico* modelling tools to be at an elementary stage for NMs [7,25,116] stressing the need to develop and provide a set of standard predictive models with defined parameters that can accurately and efficiently predict human and ecological toxicity of NM with minimal biological experimentation.

The NanoSolveIT advanced nanoinformatics approach addresses these needs by developing integrated models within a benchmarked IATA for nanosafety assessment. The innovative modelling techniques and tools will be integrated within NanoSolveIT IATA and then incorporated into a sustainable interoperable product, the NanoSolveIT e-platform. NanoSolveIT will deliver an innovative alternative testing strategy that is less reliant on animal testing for NMs RA using predominantly *in silico* derived NM descriptors to generate validated predictive models for NMs properties and adverse effects.

To accelerate implementation of the advances in nanoinformatics, engagement with regulators and policy makers is needed, to create dialog and transfer NanoSolveIT knowledge. Within NanoSolveIT a strong communication, dissemination and exploitation strategy for spreading and circulating information, managing effective communication and delivery of information, and up-skilling of relevant stakeholders enabling them to make use of and benefit from the created knowledge and resources, is implemented in parallel with the technological advances. This will ensure long-term exploitation of the NanoSolveIT IATA and outcomes, across the full range of regulatory bodies (e.g., ECHA, EFSA, SCCS, OECD, etc.). By direct communication with regulatory bodies, including risk assessors as members of the Advisory board, and using advanced communication tools to engage all relevant stakeholders NanoSolveIT aims to create common ground for exchange of information between scientists, industry, risk assessors, regulators, and policy makers. For example, several stakeholder workshops have been planned, aligning with similar nanoinformatics activities and parallel risk governance projects, to ensure that tools developed within NanoSolveIT will become utilized as an essential element for supporting industrial and regulatory nano-risk governance.

NanoSolveIT also offers the opportunity for alignment and harmonization of international activities into the ongoing and emerging activities in Europe. This will be achieved through the participation of international partners who will participate in the benchmarking of the various tools and models and the assessment of their suitability for addressing different safe-by-design, grouping and categorization, and predictive (eco)toxicity questions via the Round Robin testing. This activity also relies on meaningful engagement with international activities such as the OECD, including via the Malta Initiative, on the revision of test guidelines and development of predictive modelling approaches.

3. Summary and outlook

Significant advances have been made during the last decade in the field of nanoinformatics, which have resulted in the development of various modelling frameworks, data platforms and knowledge infrastructures, and *in silico* tools for generating meaningful hazard and RA predictions for NMs. A short analysis of the recent advances in the distinctive nanoinformatics fields has been provided here, identifying the state-of-the-art, the current gaps in

research and potential challenges to be solved by the development of innovative techniques and methods.

First, the issues of data quality, reliability and accessibility have to be addressed in a harmonized manner by ensuring that all NMs data platforms are connected and follow the same standards in order to allow easy access to curated, relevant data, under the FAIR principles. Quality and reliability assessment of data for NMs remains an important challenge and should be performed as early as possible in the data generation process, preferably at the point of data generation as promoted by the NanoCommons nanosafety research infrastructure [120].

Secondly, scientific tools for nanoinformatics have been, and continue to be, developed. The effort to collect and link the available tools under state-of-the-art platforms is the next step, closely related to the efforts to perform holistic risk governance processes for NMs. This would increase the accessibility of the tools and would allow possible validation of the available tools via benchmarking and comparison of predictions. This would provide high interoperability and facilitate reuse of data.

Addressing these challenges, NanoSolveIT is developing a broad spectrum of diverse but interlinked advanced physics-based, omics-based and data-driven (AI, deep learning) models that work synergistically exchanging inputs and outputs and which together constitute a multi-scale *in silico* IATA for evaluation of the environmental health and safety of NMs. This will be achieved through the development and use of an extraordinary breadth of integrated models ranging from NM-biomolecule and cell interactions, models for 'omics analysis and AOP generation, release and exposure models, toxicokinetics and PBPK models, and ecotoxicity prediction, all delivered within a benchmarked IATA for nanosafety assessment, to enable safe-by-design NM development and NMs Risk Governance. The NanoSolveIT IATA will be implemented as a decision support system packaged both as a standalone software and a Cloud platform.

Key impacts of the NanoSolveIT models and *in silico* IATA will be a direct reduction in the need for animal testing and a concurrent increase in regulatory, industry and consumer confidence in the predictive capacity of the nanoinformatics nanosafety models. The *in silico* IATA will support the implementation of "safe-built-by-design" approaches in industry by applying cost-effective testing platforms for exposure, hazard and RA.

Acknowledging the complexity of the challenge, complete achievement of this vision is likely to require 10 years of research and investment, however, NanoSolveIT anticipates having a first version of the concept implemented by early 2023, utilizing open access approaches where possible to allow further community development and continuous enrichment of the knowledgebase underpinning the models. Exciting times are ahead for the nanosafety Nanoinformatics community.

CCRediT authorship contribution statement

Antreas Afantitis: Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Georgia Melagraki:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Panagiotis Isigonis:** Writing - original draft. **Andreas Tsoumanis:** Writing - original draft. **Dimitra Danai Var-sou:** Writing - original draft. **Eugenia Valsami-Jones:** Writing - original draft, Writing - review & editing. **Anastasios Papadiamantis:** Writing - original draft. **Laura-Jayne A Ellis:** Writing - original draft. **Haralambos Sarimveis:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Philip Doganis:** Conceptualization, Writing - original draft, Writing - review & editing. **Pantelis Karatzas:** Writing - original draft. **Periklis Tsiros:** Writing - original draft. **Irene Liampa:** Writing - original draft. **Vla-**

dimir Lobaskin: Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Dario Greco:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing. **Angela Serra:** Writing - original draft. **Pia Anneli Sofia Kinaret:** Writing - original draft. **Laura Aliisa Saarimäki:** Writing - original draft. **Roland Grafström:** Conceptualization, Supervision, Writing - review & editing. **Pekka Kohonen:** Writing - original draft. **Penny Nymark:** Writing - original draft, Writing - review & editing. **Egon Willighagen:** Conceptualization, Supervision, Writing - review & editing. **Tomasz Puzyn:** Conceptualization, Supervision, Writing - review & editing. **Anna Rybinska-Fryca:** Writing - original draft. **Alexander Lyubartsev:** Writing - original draft. **Keld Alstrup Jensen:** Writing - original draft. **Jan Gerit Brandenburg:** Writing - review & editing. **Stephen Lofts:** Conceptualization, Writing - review & editing. **Claus Svendsen:** Conceptualization, Supervision. **Samuel Harrison:** Writing - original draft. **Dieter Maier:** Conceptualization, Writing - original draft, Writing - review & editing. **Kaido Tamm:** Conceptualization, Writing - original draft, Writing - review & editing. **Jaak Jänes:** Writing - review & editing. **Lauri Sikk:** Writing - review & editing. **Maria Dusinska:** Conceptualization, Writing - original draft, Writing - review & editing. **Eleonora Longhin:** Writing - original draft. **Elise Rundén-Pran:** Writing - original draft. **Espen Mariussen:** Writing - original draft. **Naouale El Yamani:** Writing - review & editing. **Wolfgang Unger:** Writing - original draft. **Jörg Radnik:** Writing - original draft, Writing - review & editing. **Alexander Tropsha:** Writing - original draft, Writing - review & editing. **Yoram Cohen:** Writing - original draft, Writing - review & editing. **Jerzy Leszczynski:** Writing - original draft, Writing - review & editing. **Christine Ogilvie Hendren:** Writing - original draft, Writing - review & editing. **Mark Wiesner:** Writing - original draft, Writing - review & editing. **David Winkler:** Conceptualization, Writing - original draft, Writing - review & editing. **Noriyuki Suzuki:** Writing - original draft, Writing - review & editing. **Tae Hyun Yoon:** Conceptualization, Writing - original draft, Writing - review & editing. **Jang-Sik Choi:** Writing - original draft, Writing - review & editing. **Natasha Sanabria:** Writing - original draft, Writing - review & editing. **Mary Gulumian:** Conceptualization, Writing - original draft, Writing - review & editing. **Iseult Lynch:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme via NanoSolveIT Project under grant agreement No 814572. Tae Hyun Yoon (THY) and Jang-Sik Choi (JSC) also appreciate partial support by the Korean Ministry of Science & ICT (MSIT) and the National Research Foundation (NRF) through the International Cooperative R&D Program (No. 2019K1A3A1A78112959), as a part of the European Union's Horizon 2020 NanoSolveIT Project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.02.023>.

References

- Afantitis A, Melagraki G, Tsoumanis A, Valsami-Jones E, Lynch I. A nanoinformatics decision support tool for the virtual screening of gold nanoparticle cellular association using protein corona fingerprints. *Nanotoxicology* 2018;12(10):1148–65. <https://doi.org/10.1080/17435390.2018.1504998>.
- Albanese A, Walkey CD, Olsen JB, Guo H, Emili A, Chan WCW. Secreted biomolecules alter the biological identity and cellular interactions of nanoparticles. *ACS Nano* 2014;8(6):5515–26. <https://doi.org/10.1021/nn4061012>.
- Alexander DJ, Tropsha A, Winkler DA. Beware of R2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *J Chem Inf Model* 2015;55(7):1316–22. <https://doi.org/10.1021/acs.jcim.5b00206>.
- Altrees-Williams S, Clapp R. Specific toxicity and crystallinity of α -quartz in respirable dust samples. *Am Ind Hyg Assoc J* 2002;63(3):348–53. <https://doi.org/10.1080/15428110208984724>.
- An H, Ling C, Xu M, Hu M, Wang H, Liu J, et al. Oxidative damage induced by nano-titanium dioxide in rats and mice: a systematic review and meta-analysis. *Biol Trace Elem Res* 2019;1–19. <https://doi.org/10.1007/s12011-019-01761-z>.
- Autefage H, Gentleman E, Littmann E, Hedegaard MAB, Von Erlach T, O'Donnell M, et al. Sparse feature selection methods identify unexpected global cellular response to strontium-containing materials. In: *Pro Natl Acad Sci*. <https://doi.org/10.1073/pnas.1419799112>.
- Bernauer U, Bodin L, Pieter-Jan C, Dusinska M, Janine E, Gaffet E, et al. TH REVISION "The SCCS note of Guidance for the testing of cosmetic ingredients and their safety evaluation – 10th Revision" SCCS/1602/18 – Final version. THE SCCS NOTES OF GUIDANCE FOR THE TESTING OF COSMETIC INGREDIENTS AND THEIR SAFETY EVALUATION 2018;10.
- Brandt EG, Lyubartsev AP. Systematic optimization of a force field for classical simulations of TiO_2 –water interfaces. *J Phys Chem C* 2015;119(32):18110–25. <https://doi.org/10.1021/acs.jpcc.5b02669>.
- Burden FR, Winkler DA. An optimal self-pruning neural network and nonlinear descriptor selection in QSAR. *QSAR Comb Sci* 2009;28(10):1092–7. <https://doi.org/10.1002/qsar.200810202>.
- Burello E, Worth AP. A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. *Nanotoxicology* 2011;5(2):228–35. <https://doi.org/10.3109/17435390.2010.502980>.
- Burk J, Sikk L, Burk P, Manshian BB, Soenen SJ, Scott-Fordsmand JJ, et al. Fe-Doped ZnO nanoparticle toxicity: assessment by a new generation of nanodescriptors. *Nanoscale* 2018;10(46):21985–93. <https://doi.org/10.1039/c8nr05220d>.
- Casals E, Gusta MF, Piella J, Casals G, Jiménez W, Puentes V. Intrinsic and extrinsic properties affecting innate immune responses to nanoparticles: the case of cerium oxide. *Front Immunol* 2017;8. <https://doi.org/10.3389/fimmu.2017.00970>.
- Cedervall T, Lynch I, Foy M, Berggård T, Donnelly SC, Cagney G, et al. Detailed identification of plasma proteins adsorbed on copolymer nanoparticles. *Angew Chem Int Ed* 2007;46(30):5754–6. <https://doi.org/10.1002/anie.200700465>.
- Chaudhry Q, Bouwmeester H, Hertel RF. The current risk assessment paradigm in relation to the regulation of nanotechnologies. In: A.M. Graeme A. Hodge, Diana Bowman, editor. *International handbook on regulating nanotechnologies*. Edward Elgar Publishing Ltd.; 2010. p. 124–43.
- Chen C, Waller T, Walker SL. Visualization of transport and fate of nano and micro-scale particles in porous media: modeling coupled effects of ionic strength and size. *Environ Sci Nano* 2017;4(5):1025–36. <https://doi.org/10.1039/C6EN00558F>.
- Chetwynd AJ, Lynch I. The rise of the nanomaterial metabolite (small molecule) corona, and emergence of the complete corona (in press). *ES: Nano*. In Press, n.d.
- Chetwynd AJ, Wheeler KE, Lynch I. Best practice in reporting corona studies: minimum information about nanomaterial biocorona experiments (MINBE). *Nano Today* 2019;100758. <https://doi.org/10.1016/j.nantod.2019.06.004>.
- Chomenidis C, Drakakis G, Tsiliki G, Anagnostopoulou E, Valsamis A, Droganis P, et al. Jaqpot Quattro: a novel computational web platform for modeling and analysis in nanoinformatics. *J Chem Inf Model* 2017;57(9):2161–72. <https://doi.org/10.1021/acs.jcim.7b00223>.
- Constantinou AC, Fenton N, Neil M. Integrating expert knowledge with data in Bayesian networks: preserving data-driven expectations when the expert variables remain unobserved. *Expert Syst Appl* 2016;56:197–208. <https://doi.org/10.1016/j.eswa.2016.02.050>.
- Dawson KA, Anguissola S, Lynch I. The need for *in situ* characterisation in nanosafety assessment: funded transnational access via the QNano research infrastructure. *Nanotoxicology* 2012;7(3):346–9. <https://doi.org/10.3109/17435390.2012.658096>.
- Deng ZJ, Liang M, Monteiro M, Toth I, Minchin RF. Nanoparticle-induced unfolding of fibrinogen promotes Mac-1 receptor activation and inflammation. *Nat Nanotechnol* 2011;6(1):39–44. <https://doi.org/10.1038/nnano.2010.250>.
- Ding F, Radic S, Chen R, Chen P, Geitner NK, Brown JM, et al. Direct observation of a single nanoparticle-ubiquitin corona formation. *Nanoscale* 2013;5(19):9162–9. <https://doi.org/10.1039/c3nr02147e>.
- Dobrovolskaia MA, Neun BW, Man S, Ye X, Hansen M, Patri AK, et al. Protein corona composition does not accurately predict hemato-compatibility of colloidal gold nanoparticles. *Nanomed Nanotechnol Biol Med* 2014;10(7):1453–63. <https://doi.org/10.1016/j.nano.2014.01.009>.
- Donaldson K, Poland CA. Nanotoxicity: challenging the myth of nano-specific toxicity. *Curr Opin Biotechnol* 2013;24:724–34. <https://doi.org/10.1016/j.copbio.2013.05.003>.
- EFSA. EFSA Guidance on risk assessment of the application of nanoscience and nanotechnologies in the food and feed chain: Part 1, human and animal health. 2018. <https://doi.org/10.2903/j.efsa.2018.5327>.
- EMMC. Charter for EMMC team on Validation and Characterisation Retrieved from https://emmc.info/wp-content/uploads/2015/06/Charter-Validation-and-Characterisation_20140930_AdB_accepted.pdf; 2014.

- [27] enanmapper/nanojava: Java library for descriptor calculation for (nano)materials. (n.d.). Retrieved February 18, 2020, from <https://github.com/enanmapper/nanojava>.
- [28] European Commission. THE SCCS NOTES OF GUIDANCE FOR THE TESTING OF COSMETIC INGREDIENTS Retrieved from. SCCS 2016;1564:151. http://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_190.pdf.
- [29] Fadeel B, Farcial L, Hardy B, Vázquez-Campos S, Hristozov D, Marcomini A, et al. Advanced tools for the safety assessment of nanomaterials. *Nat Nanotechnol* 2018;13(7):537–43. <https://doi.org/10.1038/s41565-018-0185-0>.
- [30] Fadeel B, Pietrousti A, Shvedova AA. Adverse effects of engineered nanomaterials: exposure, toxicology, and impact on human health (B. Fadeel, A. Pietrousti, & A. A. Shvedova, Eds.). Academic Press, 2017.
- [31] Findlay MR, Freitas DN, Mobed-Miremadi M, Wheeler KE. Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environ Sci Nano* 2018;5(1):64–71. <https://doi.org/10.1039/C7EN00466D>.
- [32] Fleischer CC, Payne CK. Secondary structure of corona proteins determines the cell surface receptors used by nanoparticles. *J Phys Chem B* 2014;118(49):14017–26. <https://doi.org/10.1021/jp502624n>.
- [33] Forest V, Hochepied J-F, Pourchez J. Importance of choosing relevant biological end points to predict nanoparticle toxicity with computational approaches for human health risk assessment. *Chem Res Toxicol* 2019;32(7):1320–6. <https://doi.org/10.1021/acs.chemrestox.9b00022>.
- [34] Fortino V, Kinaré P, Fyhrquist N, Alenius H, Greco D. A robust and accurate method for feature selection and prioritization from multi-class OMICS data. *PLoS ONE* 2014. <https://doi.org/10.1371/journal.pone.0107801>.
- [35] Fröhlich E. Role of omics techniques in the toxicity testing of nanoparticles. *J Nanobiotechnol* 2017;15(1):84. <https://doi.org/10.1186/s12951-017-0320-3>.
- [36] Gajewicz A, Jagiello K, Cronin MTD, Leszczynski J, Puzyn T. Addressing a bottle neck for regulation of nanomaterials: quantitative read-across (Nano-QRA) algorithm for cases when only limited data is available. *Environ Sci Nano* 2017;4(2):346–58. <https://doi.org/10.1039/C6EN00399K>.
- [37] Gajewicz Agnieszka. What if the number of nanotoxicity data is too small for developing predictive Nano-QSAR models? An alternative read-across based approach for filling data gaps. *Nanoscale* 2017;9(24):8435–48. <https://doi.org/10.1039/C7NR02211F>.
- [38] Gao Z, Varela JA, Groc L, Lounis B, Cognet L. Toward the suppression of cellular toxicity from single-walled carbon nanotubes. *Biomater Sci* 2016;4:230–44. <https://doi.org/10.1039/c5bm00134j>.
- [39] Gattoo MA, Naseem S, Arfat MY, Dar AM, Qasim K, Zubair S. Physicochemical properties of nanomaterials: implication in associated toxic manifestations. *Biomed Res Int* 2014;2014. <https://doi.org/10.1155/2014/498420>.
- [40] Gernand JM, Casman EA. A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. *Risk Anal* 2014;34(3):583–97. <https://doi.org/10.1111/risa.12109>.
- [41] Goede H, Christopher-de Vries Y, Kuijpers E, Fransman W. A review of workplace risk management measures for nanomaterials to mitigate inhalation and dermal exposure. *Ann Work Exposures Health* 2018;62(8):907–22. <https://doi.org/10.1093/annweh/wxy032>.
- [42] Goodman CM, McCusker CD, Yilmaz T, Rotello VM. Toxicity of gold nanoparticles functionalized with cationic and anionic side chains. *Bioconjug Chem* 2004;15(4):897–900. <https://doi.org/10.1021/bc049951i>.
- [43] Grafström RC, Nymark P, Hongisto V, Spjuth O, Ceder R, Willighagen E, et al. Toward the replacement of animal experiments through the bioinformatics-driven analysis of 'omics' data from human cell cultures. *Altern Lab Anim* 2015;43(5):325–32. <https://doi.org/10.1177/026119291504300506>.
- [44] Guo Y, Ma T, Porter AL, Huang L. Text mining of information resources to inform forecasting innovation pathways. *Technol Anal Strat Manage* 2012;24(8):843–61. <https://doi.org/10.1080/09537325.2012.715491>.
- [45] Ha MK, Trinh TX, Choi JS, Maulina D, Byun HG, Yoon TH. Toxicity classification of oxide nanomaterials: effects of data gap filling and pchem score-based screening approaches. *Sci Rep* 2018;8(1):3141. <https://doi.org/10.1038/s41598-018-21431-9>.
- [46] Haase, Klaessig. EU US Roadmap Nanoinformatics 2030, 2018. <https://doi.org/10.5281/ZENODO.1486012>.
- [47] Hadrup N, Bengtson S, Jacobsen NR, Jackson P, Nocun M, Saber AT, et al. Influence of dispersion medium on nanomaterial-induced pulmonary inflammation and DNA strand breaks: investigation of carbon black, carbon nanotubes and three titanium dioxide nanoparticles. *Mutagenesis* 2017;32(6):581–97. <https://doi.org/10.1093/mutage/gev042>.
- [48] Hastings J, Jeliazkova N, Owen G, Tsiliki G, Munteanu CR, Steinbeck C, et al. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. *J Biomed Semantics* 2015;6:10. <https://doi.org/10.1186/s13326-015-0005-5>.
- [49] Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI - the worldwide chemical structure identifier standard. *J Cheminf* 2013;5(1):7. <https://doi.org/10.1186/1758-2946-5-7>.
- [50] Hendren CO, Powers CM, Hoover MD, Harper SL. The Nanomaterial data curation initiative: a collaborative approach to assessing, evaluating, and advancing the state of the field. *Beilstein J Nanotechnol* 2015;6(1):1752–62. <https://doi.org/10.3762/bjnano.6.179>.
- [51] Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, et al. Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res* 2015;43(D1):D921–7. <https://doi.org/10.1093/nar/eku955>.
- [52] Ilves M, Kinaré PAS, Ndika J, Karisola P, Marwah V, Fortino V, et al. Surface PEGylation suppresses pulmonary effects of CuO in allergen-induced lung inflammation. *Part Fibre Toxicol* 2019;16(1):28. <https://doi.org/10.1186/s12989-019-0309-1>.
- [53] Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. *PNAS* 2010;107(33):14621–6. <https://doi.org/10.1073/pnas.1000138107>.
- [54] Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, et al. FAIR principles: interpretations and implementation considerations. *Data Intelligence* 2020;2(1–2):10–29. https://doi.org/10.1162/dint_r.00024.
- [55] Jagiello K, Chomicz B, Avramopoulos A, Gajewicz A, Mikolajczyk A, Bonifassi P, et al. Size-dependent electronic properties of nanomaterials: How this novel class of nanodescriptors supposed to be calculated?. *Struct Chem* 2017;28(3):635–43. <https://doi.org/10.1007/s11224-016-0838-2>.
- [56] Jagiello K, Grzonkowska M, Swirog M, Ahmed L, Rasulev B, Avramopoulos A, et al. Advantages and limitations of classic and 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene derivatives. *J Nanopart Res* 2016;18(9):256. <https://doi.org/10.1007/s11051-016-3564-1>.
- [57] Jensen A, Dal Maso M, Koivisto A, Belut E, Meyer-Plath A, Van Tongeren M, et al. Comparison of geometrical layouts for a multi-box aerosol model from a single-chamber dispersion study. *Environments* 2018;5(5):52. <https://doi.org/10.3390/environments5050052>.
- [58] Kar S, Gajewicz A, Puzyn T, Roy K, Leszczynski J. Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: a mechanistic QSTR approach. *Ecotoxicol Environ Saf* 2014. <https://doi.org/10.1016/j.ecoenv.2014.05.026>.
- [59] Karakoti AS, Hench LL, Seal S. The potential toxicity of nanomaterials - the role of surfaces. *JOM* 2006;58:77–82. <https://doi.org/10.1007/s11837-006-0147-0>.
- [60] Karcher S, Willighagen EL, Rumble J, Ehrhart F, Evelo CT, Fritts M, et al. Integration among databases and data sets to support productive nanotechnology: challenges and recommendations. *NanoImpact* 2018;9:85–101. <https://doi.org/10.1016/j.nimpact.2017.11.002>.
- [61] Karlsson HL, Gustafsson J, Cronholm P, Möller L. Size-dependent toxicity of metal oxide particles—a comparison between nano- and micrometer size. *Toxicol Lett* 2009;188(2):112–8. <https://doi.org/10.1016/j.toxlet.2009.03.014>.
- [62] Kharazian B, Lohse SE, Ghasemi F, Raoufi M, Saei AA, Hashemi F, et al. Bare surface of gold nanoparticle induces inflammation through unfolding of plasma fibrinogen. *Sci Rep* 2018;8(1):1–9. <https://doi.org/10.1038/s41598-018-30915-7>.
- [63] Kinaré P, Ilves M, Fortino V, Rydman E, Karisola P, Lähde A, et al. Inhalation and oropharyngeal aspiration exposure to rod-like carbon nanotubes induce similar airway inflammation and biological responses in mouse lungs. *ACS Nano* 2017;11(1):291–303. <https://doi.org/10.1021/acsnano.6b05652>.
- [64] Kinaré P, Marwah V, Fortino V, Ilves M, Wolff H, Ruokolainen L, et al. Network analysis reveals similar transcriptomic responses to intrinsic properties of carbon nanomaterials *in vitro* and *in vivo*. *ACS Nano* 2017;11(4):3786–96. <https://doi.org/10.1021/acsnano.6b08650>.
- [65] Klimisch H-J, Andreae M, Tillmann U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharm* 1997;25(1):1–5. <https://doi.org/10.1006/RTPH.1996.1076>.
- [66] Kohonen P, Ceder R, Smit I, Hongisto V, Myatt G, Hardy B, et al. Cancer biology, toxicology and alternative methods development go hand-in-hand. *Basic Clin Pharmacol Toxicol* 2014;115(1):50–8. <https://doi.org/10.1111/bcpt.12257>.
- [67] Kohonen P, Parkkinen JA, Willighagen EL, Ceder R, Wennerberg K, Kaski S, et al. A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nat Commun* 2017;8(1):15932. <https://doi.org/10.1038/ncomms15932>.
- [68] Koivisto AJ, Aromaa M, Koponen IK, Fransman W, Jensen KA, Mäkelä JM, et al. Workplace performance of a loose-fitting powered air purifying respirator during nanoparticle synthesis. *J Nanopart Res* 2015;17(4):1–10. <https://doi.org/10.1007/s11051-015-2990-9>.
- [69] Kuz'min VE, Artemenko AG, Muratov EN. Hierarchical QSAR technology based on the Simplex representation of molecular structure. *J Comput Aided Mol Des* 2008;22(6–7):403–21. <https://doi.org/10.1007/s10822-008-9179-6>.
- [70] Labib S, Williams A, Yauk CL, Nikota JK, Wallin H, Vogel U, et al. Nano-risk science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Part Fibre Toxicol* 2015;13(1):15. <https://doi.org/10.1186/s12989-016-0125-9>.
- [71] Labouta HI, Asgarian N, Rinker K, Cramb DT. Meta-analysis of nanoparticle cytotoxicity via data-mining the literature. *ACS Nano* 2019;acs.nano.8b07562. <https://doi.org/10.1021/acsnano.8b07562>.
- [72] Latour RA. Perspectives on the simulation of protein-surface interactions using empirical force field methods. *Colloids Surf, B* 2014;124:25–37. <https://doi.org/10.1016/j.colsurfb.2014.06.050>.
- [73] Le T, Epa VC, Burden FR, Winkler DA. Quantitative structure-property relationship modeling of diverse materials properties. *Chem Rev* 2012;112(5):2889–919. <https://doi.org/10.1021/cr200066h>.

- [74] Lee YK, Choi EJ, Webster TJ, Kim SH, Khang D. Effect of the protein corona on nanoparticles for modulating cytotoxicity and immunotoxicity. *Int J Nanomed* 2014;10:97–113. <https://doi.org/10.2147/IJN.S72998>.
- [75] Legradi JB, Di Paolo C, Kraak MHS, van der Geest HG, Schymanski EL, Williams AJ, et al. An ecotoxicological view on neurotoxicity assessment. *Environ Sci Eur* 2018;30(1):46. <https://doi.org/10.1186/s12302-018-0173-x>.
- [76] Leong HS, Butler KS, Brinker KS, Azzawi M, Conlan S, Dufès C, et al. On the issue of transparency and reproducibility in nanomedicine. *Nat Nanotechnol* 2019;14(7):629–35. <https://doi.org/10.1038/s41565-019-0496-9>.
- [77] Leroueil PR, Berry SA, Duthie K, Han G, Rotello VM, McNerny DQ, et al. Wide varieties of cationic nanoparticles induce defects in supported lipid bilayers. *Nano Lett* 2008;8(2):420–4. <https://doi.org/10.1021/nl0722929>.
- [78] Lewinski NA, McInnes BT. Using natural language processing techniques to inform research on nanotechnology. *Beilstein J Nanotechnol* 2015;6(1):1439–49. <https://doi.org/10.3762/bjnano.6.149>.
- [79] Li N, Sioutas C, Cho A, Schmitz D, Misra C, Sempf J, et al. Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Environ Health Perspect* 2003;111(4):455–60. <https://doi.org/10.1289/ehp.6000>.
- [80] Liu R, Jiang W, Walkey CD, Chan WCW, Cohen Y. Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale* 2015;7(21):9664–75. <https://doi.org/10.1039/C5NR01537E>.
- [81] Lopez H, Brandt EG, Mirzoev A, Zhurkin D, Lyubartsev A, Lobaskin V. Multiscale modelling of Bionano Interface 2017. https://doi.org/10.1007/978-3-319-47754-1_7.
- [82] Lopez H, Lobaskin V. Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. *J Chem Phys* 2015;143(24):. <https://doi.org/10.1063/1.4936908>243138.
- [83] Lubinski L, Urbaszek P, Gajewicz A, Cronin MTD, Enoch SJ, Madden JC, et al. Evaluation criteria for the quality of published experimental data on nanomaterials and their usefulness for QSAR modelling. *SAR QSAR Environ Res* 2013;24(12):995–1008. <https://doi.org/10.1080/1062936X.2013.840679>.
- [84] Lundqvist M, Stigler J, Cedervall T, Berggård T, Flanagan MB, Lynch I, et al. The evolution of the protein corona around nanoparticles: a test study. *ACS Nano* 2011;5(9):7503–9. <https://doi.org/10.1021/nn202458g>.
- [85] Lundqvist M, Stigler J, Elia G, Lynch I, Cedervall T, Dawson KA. Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. *Proc Natl Acad Sci* 2008;105(38):14265–70. <https://doi.org/10.1073/PNAS.0805135105>.
- [86] Lynch I. Are there generic mechanisms governing interactions between nanoparticles and cells? Epitope mapping the outer layer of the protein-material interface. *Physica A* 2007;373:511–20. <https://doi.org/10.1016/j.physa.2006.06.008>.
- [87] Lynch I, Afantitis A, Leonis G, Melagraki G, Valsami-Jones E. Strategy for identification of nanomaterials' critical properties linked to biological impacts: interlinking of experimental and computational approaches. In *Advances in QSAR Modeling*, 2017, pp. 385–424. https://doi.org/10.1007/978-3-319-56850-8_10.
- [88] Lynch I, Dawson KA, Linse S. Detecting cryptic epitopes created by nanoparticles. *Science's STKE : Signal Transduction Knowledge Environment* 2006;2006(327):pe14. <https://doi.org/10.1126/stke.3272006pe14>.
- [89] Lynch I, Weiss C, Valsami-Jones E. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. *Nano Today* 2014;9(3):266–70. <https://doi.org/10.1016/j.nantod.2014.05.001>.
- [90] Manganelli S, Leone C, Toropov AA, Toropova AP, Benfenati E. QSAR model for predicting cell viability of human embryonic kidney cells exposed to SiO₂ nanoparticles. *Chemosphere* 2016;144:995–1001. <https://doi.org/10.1016/j.chemosphere.2015.09.086>.
- [91] Manshian BB, Pokhrel S, Himmelfreich U, Tämml K, Sikk L, Fernández A, et al. In silico design of optimal dissolution kinetics of Fe-doped ZnO nanoparticles results in cancer-specific toxicity in a preclinical rodent model. *Adv Healthcare Mater* 2017;6(9):1601379. <https://doi.org/10.1002/adhm.201601379>.
- [92] Marchese Robinson RL, Lynch I, Peijnenburg W, Rumble J, Klaessig F, Marquardt C, et al. How should the completeness and quality of curated nanomaterial data be evaluated?. *Nanoscale* 2016;8(19):9919–43. <https://doi.org/10.1039/C5NR08944A>.
- [93] Marwah VS, Kinaret PAS, Serra A, Scala G, Lauerma A, Fortino V, et al. InFORM: inference of network response modules. *Bioinformatics* (Oxford, England) 2018;34(12):2136–8. <https://doi.org/10.1093/bioinformatics/bty063>.
- [94] Marwah VS, Scala G, Kinaret PAS, Serra A, Alenius H, Fortino V, et al. eUTOPIA: solUTion for omics data preprocessing and analysis. *Source Code Biol Med* 2019;14(1):1. <https://doi.org/10.1186/s13029-019-0071-7>.
- [95] Matatiele P, Mochaki L, Southon B, Dabula B, Poongavanum P, Kgarebe B. Environmental and biological monitoring in the workplace: a 10-year South African retrospective analysis. *AAS Open Research* 2019;1:20. <https://doi.org/10.12688/aasopenres.12882.2>.
- [96] McNaught AD, Wilkinson A. Compendium of chemical terminology: IUPAC recommendations, 1997. Retrieved from <http://www.old.iupac.org/publications/books/author/mcnaught.html>.
- [97] Melagraki G, Afantitis A. A risk assessment tool for the virtual screening of metal oxide nanoparticles through Enalos InSilicoNano Platform. *Curr Top Med Chem* 2015;15:1827–36.
- [98] Melagraki Georgia, Afantitis A. Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. *RSC Adv* 2014;4(92):50713–25. <https://doi.org/10.1039/C4RA07756C>.
- [99] Mikolajczyk A, Gajewicz A, Rasulev B, Schaeublin N, Maurer-Gardner E, Hussain S, et al. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. *Chem Mater* 2015;27(7):2400–7. <https://doi.org/10.1021/cm504406a>.
- [100] Morris SA, Gaheen S, Lijowski M, Heiskanen M, Klemm J. Experiences in supporting the structured collection of cancer nanotechnology data using caNanoLab. *Beilstein J Nanotechnol* 2015;6(1):1580–93. <https://doi.org/10.3762/bjnano.6.161>.
- [101] Mukherjee SP, Bondarenko O, Kohonen P, Andón FT, Brzicová T, Gessner I, et al. Macrophage sensing of single-walled carbon nanotubes via Toll-like receptors. *Sci Rep* 2018;8(1):1115. <https://doi.org/10.1038/s41598-018-19521-9>.
- [102] Mühlhopt S, Diabaté S, Dilger M, Adelhelm C, Anderlohr C, Bergfeldt T, et al. Characterization of nanoparticle batch-to-batch variability. *Nanomaterials* 2018;8(5):311. <https://doi.org/10.3390/nano8050311>.
- [103] Müller KH, Kulkarni J, Motskin M, Goode A, Winship P, Skepper JN, et al. pH-dependent toxicity of high aspect ratio ZnO nanowires in macrophages due to intracellular dissolution. *ACS Nano* 2010;4(11):6767–79. <https://doi.org/10.1021/nn101192z>.
- [104] Murphy F, Sheehan B, Mullins M, Bouwmeester H, Marvin HJP, Bouzembrak Y, et al. A tractable method for measuring nanomaterial risk using bayesian networks. *Nanoscale Res Lett* 2016;11(1). <https://doi.org/10.1186/s11671-016-1724-y>.
- [105] Nagai H, Okazaki Y, Chew SH, Misawa N, Yamashita Y, Akatsuka S, et al. Diameter and rigidity of multiwalled carbon nanotubes are critical factors in mesothelial injury and carcinogenesis. *Proc Natl Acad Sci* 2011;108(49):E1330–8. <https://doi.org/10.1073/PNAS.1110013108>.
- [106] Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, et al. Drug repositioning: a machine-learning approach through data integration. *J Cheminf* 2013;5(1):30. <https://doi.org/10.1186/1758-2946-5-30>.
- [107] Nikota J, Williams A, Yauk CL, Wallin H, Vogel U, Halappanavar S. Meta-analysis of transcriptomic responses as a means to identify pulmonary disease outcomes for engineered nanomaterials. *Part Fibre Toxicol* 2015;13(1):25. <https://doi.org/10.1186/s12989-016-0137-5>.
- [108] NovaMechanics Ltd. Enalos Cloud Platform Retrieved November 11, 2019, from <http://enaloscloud.novamechanics.com/>; 2019.
- [109] Nymark P, Bakker M, Dekkers S, Franken R, Fransman W, García-Bilbao A, et al. Toward rigorous materials production: new approach methodologies have extensive potential to improve current safety assessment practices. *Small* 2020;16(6):. <https://doi.org/10.1002/smll.201904749>e1904749.
- [110] Nymark P, Kohonen P, Hongisto V, Grafström RC. Toxic and genomic influences of inhaled nanomaterials as a basis for predicting adverse outcome. *Ann Am Thorac Soc* 2018;15(Supplement_2):S91–7. <https://doi.org/10.1513/AnnalsATS.201706-478MG>.
- [111] Nymark P, Rieswijk L, Ehrhart F, Jeliakova N, Tsiliki G, Sarimveis H, et al. A data fusion pipeline for generating and enriching adverse outcome pathway descriptions. *Toxicol Sci* 2018;162(1):264–75. <https://doi.org/10.1093/toxsci/kfx252>.
- [112] Oberdörster E. Manufactured nanomaterials (Fullerenes, C₆₀) induce oxidative stress in the brain of juvenile largemouth bass. *Environ Health Perspect* 2004;112(10):1058–62. <https://doi.org/10.1289/ehp.7021>.
- [113] Odzimek K, Ushizima D, Oberbek P, Kurzydowski KJ, Puzyr T, Haranczyk M. Scanning electron microscopy image representativeness: morphological data on nanoparticles. *J Microsc* 2017;265(1):34–50. <https://doi.org/10.1111/jmi.12461>.
- [114] OECD. No. 69 - Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. Paris, 2007.
- [115] OECD. No 25 - Guidance Manual for the Testing of Manufactured Nanomaterials: OECD Sponsorship Programme. Paris: First Revision; 2010.
- [116] OECD. ALTERNATIVE TESTING STRATEGIES IN RISK ASSESSMENT OF MANUFACTURED NANOMATERIALS: CURRENT STATE OF KNOWLEDGE AND RESEARCH NEEDS TO ADVANCE THEIR USE Retrieved from <https://one.oecd.org/document/ENV/JM/MONO%282016%2963/en/pdf>; 2017.
- [117] OECD. DEVELOPMENTS IN DELEGATIONS ON THE SAFETY OF MANUFACTURED NANOMATERIALS –TOUR DE TABLE. Series on the Safety of Manufactured Nanomaterials No. 89, 2019.
- [118] Oh E, Liu R, Nel A, Gemill KB, Bilal M, Cohen Y, et al. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. *Nat Nanotechnol* 2016;11(5):479–86. <https://doi.org/10.1038/nnano.2015.338>.
- [119] Oksel C, Ma CY, Liu JJ, Wilkins T, Wang XZ. (Q)SAR modelling of nanomaterial toxicity: a critical review. *Particuology* 2015;21:1–19. <https://doi.org/10.1016/j.PARTIC.2014.12.001>.
- [120] Papadiamantis A, Farcas L, Willighagen E, Lynch I, Exner T. D10.1 Initial draft of data management plan (Open data pilot) v 2.0, 2019. <https://doi.org/10.5281/ZENODO.3404018>.
- [121] Park E-J, Park Y-K, Park K-S. Acute toxicity and tissue distribution of cerium oxide nanoparticles by a single oral administration in rats. *Toxicol Res* 2009;25(2):79–84. <https://doi.org/10.5487/TR.2009.25.2.079>.
- [122] Perelshtein I, Lipovsky A, Perkash N, Gedanken A, Moschini E, Mantecca P. The influence of the crystalline nature of nano-metal oxides on their antibacterial and toxicity properties. *Nano Res* 2015;8(2):695–707. <https://doi.org/10.1007/s12274-014-0553-5>.

- [123] Petrova T, Rasulev BF, Toropov AA, Leszczynska D, Leszczynski J. Improved model for fullerene C60 solubility in organic solvents based on quantum-chemical and topological descriptors. *J Nanopart Res* 2011;13(8):3235–47. <https://doi.org/10.1007/s11051-011-0238-x>.
- [124] Podila R, Brown JM. Toxicity of engineered nanomaterials: a physicochemical perspective. *J Biochem Mol Toxicol* 2013;27(1):50–5. <https://doi.org/10.1002/jbt.21442>.
- [125] Poland CA, Duffin R, Kinloch I, Maynard A, Wallace WAH, Seaton A, et al. Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study. *Nat Nanotechnol* 2008;3(7):423–8. <https://doi.org/10.1038/nnano.2008.111>.
- [126] Poulsen SS, Jackson P, Kling K, Knudsen KB, Skaug V, Kyjovska ZO, et al. Multi-walled carbon nanotube physicochemical properties predict pulmonary inflammation and genotoxicity. *Nanotoxicology* 2016;10(9):1263–75. <https://doi.org/10.1080/17435390.2016.1202351>.
- [127] Poulsen SS, Knudsen KB, Jackson P, Weydahl IEK, Saber AT, Wallin H, et al. Multi-walled carbon nanotube-physicochemical properties predict the systemic acute phase response following pulmonary exposure in mice. *PLoS ONE* 2017;12(4):. <https://doi.org/10.1371/journal.pone.0174167>.
- [128] Poulsen SS, Saber AT, Williams A, Andersen O, Købler C, Atluri R, et al. MWNTs of different physicochemical properties cause similar inflammatory responses, but differences in transcriptional and histological markers of fibrosis in mouse lungs. *Toxicol Appl Pharmacol* 2015;284(1):16–32. <https://doi.org/10.1016/j.taap.2014.12.011>.
- [129] Power D, Rouse I, Poggio S, Brandt E, Lopez H, Lyubartsev A, et al. A multiscale model of protein adsorption on a nanoparticle surface. *Modell Simul Mater Sci Eng* 2019;27(8):84003. <https://doi.org/10.1088/1361-651X/ab3b6e>.
- [130] Powers CM, Mills KA, Morris SA, Klaessig F, Gaheen S, Lewinski N, et al. Nanocuration workflows: establishing best practices for identifying, inputting, and sharing data to inform decisions on nanomaterials. *Beilstein J Nanotechnol* 2015;6(1):1860–71. <https://doi.org/10.3762/bjnano.6.189>.
- [131] Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat Nanotechnol* 2011;6(3):175–8. <https://doi.org/10.1038/nnano.2011.10>.
- [132] Rahman L, Jacobsen NR, Aziz SA, Wu D, Williams A, Yauk CL, et al. Multi-walled carbon nanotube-induced genotoxic, inflammatory and pro-fibrotic responses in mice: investigating the mechanisms of pulmonary carcinogenesis. *Mutation Res/Genetic Toxicol Environ Mutagenesis* 2017;823:28–44. <https://doi.org/10.1016/j.mrgentox.2017.08.005>.
- [133] Ramaiahgari SC, Auerbach SS, Saddler TO, Rice JR, Dunlap PE, Sipes NS, et al. The power of resolution: contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. *Toxicol Sci* 2019;169(2):553–66. <https://doi.org/10.1093/toxsci/kfz065>.
- [134] Randić M. Novel graph theoretical approach to heteroatoms in quantitative structure–activity relationships. *Chemometr Intell Lab Syst* 1991;10(1–2):213–27. [https://doi.org/10.1016/0169-7439\(91\)80051-Q](https://doi.org/10.1016/0169-7439(91)80051-Q).
- [135] Reinsch BC, Levard C, Li Z, Ma R, Wise A, Gregory KB, et al. Sulfidation of silver nanoparticles decreases *Escherichia coli* growth inhibition. *Environ Sci Technol* 2012;46(13):6992–7000. <https://doi.org/10.1021/es203732x>.
- [136] Rivera Gil P, Oberdörster G, Elder A, Puentes V, Parak WJ. Correlating physicochemical with toxicological properties of nanoparticles: the present and the future. *ACS Nano* 2010;4(10):5527–31. <https://doi.org/10.1021/nn1025687>.
- [137] Roebben G, Ramirez-Garcia S, Hackley VA, Roesslein M, Klaessig F, Kestens V, et al. Interlaboratory comparison of size and surface charge measurements on nanoparticles prior to biological impact assessment. *J Nanopart Res* 2011;13(7):2675–87. <https://doi.org/10.1007/s11051-011-0423-y>.
- [138] Rojas-Guzmán C, Kramer MA. GALGO: a genetic algorithm decision support tool for complex uncertain systems modeled with bayesian belief networks. *Uncertainty Artificial Intelligence* 1993;368–375. <https://doi.org/10.1016/B978-1-4832-1451-1.50049-4>.
- [139] Römer I, Briffa SM, Arroyo Rojas Dasilva Y, Hapiuk D, Trouillet V, Palmer RE, Valsami-Jones E. Impact of particle size, oxidation state and capping agent of different cerium dioxide nanoparticles on the phosphate-induced transformations at different pH and concentration. *PLOS ONE*, 14(6), e0217483, 2019. <https://doi.org/10.1371/journal.pone.0217483>.
- [140] Russell W, Burch R, Hume C. *The principles of humane experimental technique*. London: Methuen Publishing Ltd; 1959.
- [141] Rydman EM, Ilves M, Koivisto AJ, Kinaret PAS, Fortino V, Savinko TS, et al. Inhalation of rod-like carbon nanotubes causes unconventional allergic airway inflammation. *Part Fibre Toxicol* 2014;11(1):48. <https://doi.org/10.1186/s12989-014-0048-2>.
- [142] Sayes C, Ivanov I. Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. *Risk Anal* 2010;30(11):1723–34. <https://doi.org/10.1111/j.1539-6924.2010.01438.x>.
- [143] Scala G, Kinaret P, Marwah V, Sund J, Fortino V, Greco D. Multi-omics analysis of ten carbon nanomaterials effects highlights cell type specific patterns of molecular regulation and adaptation. *Nanolimpact* 2018;11:99–108. <https://doi.org/10.1016/j.nimpact.2018.05.003>.
- [144] Scala G, Marwah V, Kinaret P, Sund J, Fortino V, Greco D. Integration of genome-wide mRNA and miRNA expression, and DNA methylation data of three cell lines exposed to ten carbon nanomaterials. *Data in Brief* 2018;19:1046–57. <https://doi.org/10.1016/j.dib.2018.05.107>.
- [145] Serra A, Letunic I, Fortino V, Handy RD, Fadeel B, Tagliaferri R, et al. INSIDE NANO: a systems biology framework to contextualize the mechanism-of-action of engineered nanomaterials. *Sci Rep* 2019;9(1):179. <https://doi.org/10.1038/s41598-018-37411-y>.
- [146] Serra A, Önlü S, Coretto P, Greco D. An integrated quantitative structure and mechanism of action-activity relationship model of human serum albumin binding. *J Cheminf* 2019;11(1):38. <https://doi.org/10.1186/s13321-019-0359-2>.
- [147] Serra A, Önlü S, Festa P, Fortino V, Greco D. MaNGA: a novel multi-niche multi-objective genetic algorithm for QSAR modelling. *Bioinformatics* 2019. <https://doi.org/10.1093/bioinformatics/btz521>.
- [148] Sigurdsson JH, Walls LA, Quigley JL. Bayesian belief nets for managing expert judgement and modelling reliability. *Qual Reliab Eng Int* 2001;17(3):181–90. <https://doi.org/10.1002/qre.410>.
- [149] Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 2019;35(17):3055–62. <https://doi.org/10.1093/bioinformatics/bty1054>.
- [150] Sizochenko N, Rasulev B, Gajewicz A, Kuz'min V, Puzyn T, Leszczynski J. From basic physics to mechanisms of toxicity: the “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. *Nanoscale* 2014;6(22):13986–93. <https://doi.org/10.1039/C4NR03487B>.
- [151] Steinhäuser KG, Sayre PG. Reliability of methods and data for regulatory assessment of nanomaterial risks. *Nanolimpact* 2017;7:66–74. <https://doi.org/10.1016/j.nimpact.2017.06.001>.
- [152] Subramanian A, Narayan R, Corseolo SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;171(6):1437–1452.e17. <https://doi.org/10.1016/j.cell.2017.10.049>.
- [153] Sun B, Pokhrel S, Dunphy DR, Zhang H, Ji Z, Wang X, et al. Reduction of acute inflammatory effects of fumed silica nanoparticles in the lung by adjusting silanol display through calcination and metal doping. *ACS Nano* 2015;9(9):9357–72. <https://doi.org/10.1021/acs.nano.5b03443>.
- [154] Swedish Chemicals Agency. Impact Assessment of Further Regulation of Nanomaterials at a European Level. Impact Assessment of Further Regulation of Nanomaterials at a European Level, 2015.
- [155] Tamm K, Sikk L, Burk J, Rallo R, Pokhrel S, Mädlar L, et al. No Title. *Nanoscale* 2016;8(36):16243–50.
- [156] Tamm K, Sikk L, Burk J, Rallo R, Pokhrel S, Mädlar L, et al. Parametrization of nanoparticles: development of full-particle nanodescriptors. *Nanoscale* 2016;8(36):16243–50. <https://doi.org/10.1039/C6NR04376C>.
- [157] Tantra R, Okseel C, Puzyn T, Wang J, Robinson KN, Wang XZ, et al. Nano(Q)SAR: challenges, pitfalls and perspectives. *Nanotoxicology* 2015;9(5):636–42. <https://doi.org/10.3109/17435390.2014.952698>.
- [158] Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, et al. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol Sci* 2019;169(2):317–32. <https://doi.org/10.1093/toxsci/kfz058>.
- [159] Toropov AA, Leszczynska D, Leszczynski J. Predicting thermal conductivity of nanomaterials by correlation weighting technological attributes codes. *Mater Lett* 2007;61(26):4777–80. <https://doi.org/10.1016/j.matlet.2007.03.026>.
- [160] Toropov AA, Leszczynska D, Leszczynski J. Predicting water solubility and octanol water partition coefficient for carbon nanotubes based on the chiral vector. *Comput Biol Chem* 2007;31(2):127–8. <https://doi.org/10.1016/j.compbiolchem.2007.02.002>.
- [161] Toropov AA, Rasulev BF, Leszczynska D, Leszczynski J. Multiplicative SMILES-based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. *Chem Phys Lett* 2008;457(4–6):332–6. <https://doi.org/10.1016/j.cplett.2008.04.013>.
- [162] Toropov AA, Toropova AP, Benfenati E, Gini G, Puzyn T, Leszczynska D, et al. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli*. *Chemosphere* 2012. <https://doi.org/10.1016/j.chemosphere.2012.05.077>.
- [163] Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J. Additive InChI-based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. *J Math Chem* 2009;46(4):1232–51. <https://doi.org/10.1007/s10910-008-9514-0>.
- [164] Toropov AA, Toropova AP, Puzyn T, Benfenati E, Gini G, Leszczynska D, et al. QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere* 2013;92(1):31–7. <https://doi.org/10.1016/j.chemosphere.2013.03.012>.
- [165] Toropov A, Sizochenko N, Toropova A, Leszczynski J. Towards the development of global nano-quantitative structure-property relationship models: zeta potentials of metal oxide nanoparticles. *Nanomaterials* 2018;8(4):243. <https://doi.org/10.3390/nano8040243>.
- [166] Toropova AP, Toropov AA, Rallo R, Leszczynska D, Leszczynski J. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions. *Ecotoxicol Environ Saf* 2015;112:39–45. <https://doi.org/10.1016/j.ecoenv.2014.10.003>.
- [167] Trevino V, Falciani F. GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* 2006;22(9):1154–6. <https://doi.org/10.1093/bioinformatics/btl074>.
- [168] Trinh TX, Ha MK, Choi JS, Byun HG, Yoon TH. Curation of datasets, assessment of their quality and completeness, and nanoSAR classification model development for metallic nanoparticles. *Environ Sci: Nano*, 2018. <https://doi.org/10.1039/C8EN00061A>.
- [169] Tsiliki G, Munteanu CR, Seoane JA, Fernandez-Lozano C, Sarimveis H, Willighagen EL. RRegrs: an R package for computer-aided model selection

- with multiple regression models. *J Cheminf* 2015;7(1):46. <https://doi.org/10.1186/s13321-015-0094-2>.
- [170] Varsou D-D, Afantitis A, Melagraki G, Sarimveis H. Read-across predictions of nanoparticle hazard endpoints: a mathematical optimization approach. *Nanoscale Adv* 2019;1(9):3485–98. <https://doi.org/10.1039/C9NA00242A>.
- [171] Varsou D-D, Afantitis A, Tsoumanis A, Papadiamantis A, Valsami-Jones E, Lynch I, et al. Zeta potential read-across model utilising nanodescriptors extracted via the nanoxtract image analysis tool available on the enalos nanoinformatics cloud platform. *SMALL* 2020. <https://doi.org/10.1002/sml.201906588>.
- [172] Varsou DD, Afantitis A, Tsoumanis A, Melagraki G, Sarimveis H, Valsami-Jones E, et al. A safe-by-design tool for functionalised nanomaterials through the Enalos Nanoinformatics Cloud platform. *Nanoscale Adv* 2019;1(2):706–18. <https://doi.org/10.1039/c8na00142a>.
- [173] Varsou DD, Nikolakopoulos S, Tsoumanis A, Melagraki G, Afantitis A. ENALOS + KNIME nodes: New cheminformatics tools for drug discovery. In *Methods in Molecular Biology* (Vol. 1824, pp. 113–138), 2018. https://doi.org/10.1007/978-1-4939-8630-9_7.
- [174] Varsou DD, Tsoumanis A, Afantitis A, Melagraki G. Enalos cloud platform: Nanoinformatics and cheminformatics tools. In *Methods in Pharmacology and Toxicology* (pp. 789–800), 2020. https://doi.org/10.1007/978-1-0716-0150-1_31.
- [175] Vo E, Zhuang Z, Horvatin M, Liu Y, He X, Rengasamy S. Respirator performance against nanoparticles under simulated workplace activities. *Ann Occupational Hygiene* 2015;59(8):1012–21. <https://doi.org/10.1093/annhvg/mev042>.
- [176] Walkey CD, Chan WCW. Understanding and controlling the interaction of nanomaterials with proteins in a physiological environment. *Chem Soc Rev* 2012;41(7):2780–99. <https://doi.org/10.1039/C1CS15233F>.
- [177] Walkey CD, Olsen JB, Song F, Liu R, Guo H, Olsen DWH, et al. Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. *ACS Nano* 2014;8(3):2439–55. <https://doi.org/10.1021/nn406018g>.
- [178] Wang A, Vangala K, Vo T, Zhang D, Fitzkee NC. A Three-Step Model for protein-gold nanoparticle adsorption. *J Phys Chem C* 2014;118(15):8134–42. <https://doi.org/10.1021/jp411543v>.
- [179] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model* 1988;28(1):31–6. <https://doi.org/10.1021/ci00057a005>.
- [180] Westmeier D, Chen C, Stauber RH, Docter D. June). The bio-corona and its impact on nanomaterial toxicity. *Eur J Nanomed* 2015;7:153–68. <https://doi.org/10.1515/einm-2015-0018>.
- [181] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Mons B. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;3(1):. <https://doi.org/10.1038/sdata.2016.18160018>.
- [182] Winkler DA, Burden FR, Yan B, Weissleder R, Tassa C, Shaw S, et al. Modelling and predicting the biological effects of nanomaterials. *SAR QSAR Environ Res* 2014;25(2):161–72. <https://doi.org/10.1080/1062936X.2013.874367>.
- [183] Winkler David A. Recent advances, and unresolved issues, in the application of computational modelling to the prediction of the biological effects of nanomaterials. *Toxicol Appl Pharmacol* 2016;299:96–100. <https://doi.org/10.1016/j.taap.2015.12.016>.
- [184] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucl Acids Res* 46(D1), D1074–D1082, 2018. <https://doi.org/10.1093/nar/gkx1037>.
- [185] Wittwehr C, Aladjov H, Ankley G, Byrne HJ, de Knecht J, Heinze E, et al. How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicol Sci* 2017;155(2):326–36. <https://doi.org/10.1093/toxsci/kfw207>.
- [186] Xia X-R, Monteiro-Riviere NA, Riviere JE. An index for characterization of nanomaterials in biological systems. *Nat Nanotechnol* 2010;5(9):671–5. <https://doi.org/10.1038/nnano.2010.164>.
- [187] Yan X, Sedykh A, Wang W, Zhao X, Yan B, Zhu H. In silico profiling nanoparticles: predictive nanomodeling using universal nanodescriptors and various machine learning approaches. *Nanoscale* 2019;11(17):8352–62. <https://doi.org/10.1039/c9nr00844f>.
- [188] Zhang H, Dunphy DR, Jiang X, Meng H, Sun B, Tarn D, et al. Processing pathway dependence of amorphous silica nanoparticle toxicity: colloidal vs pyrolytic. *J Am Chem Soc* 2012;134(38):15790–804. <https://doi.org/10.1021/ja304907c>.
- [189] Zhang H, Ji Z, Xia T, Meng H, Low-Kam C, Liu R, et al. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano* 2012;6(5):4349–68. <https://doi.org/10.1021/nn3010087>.
- [190] Zhang H, Pokhrel S, Ji Z, Meng H, Wang X, Lin S, et al. PdO doping tunes band-gap energy levels as well as oxidative stress responses to a Co3O4 p-type semiconductor in cells and the lung. *J Am Chem Soc* 2014;136(17):6406–20. <https://doi.org/10.1021/ja501699e>.
- [191] Zhao Y, Wang B, Feng W, Bai C. Nanotoxicology: toxicological and biological activities of nanomaterials Retrieved from. *Nanosci Nanotechnol* 2012:1–53. <http://www.eolss.net/sample-chapters/c05/e6-152-35-00.pdf>.
- [192] Zhernovkov V, Santra T, Cassidy H, Rukhlenko O, Matallanas D, Krstic A, et al. An integrative computational approach for a prioritization of key transcription regulators associated with nanomaterial-induced toxicity. *Toxicol Sci* 2019;171(2):303–14. <https://doi.org/10.1093/toxsci/kfz151>.