NanoCommons
Nano-Knowledge Community

The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

Grant Agreement No 731032

Deliverable Report 6.1

Deliverable	D6.1 A workflow and checklist for experimental design and informatics workflow for risk assessment for use in WP9
Work Package	WP6 - JRA4 - Tool integration for risk assessment
Delivery date	M18 - 30 June 2019
Lead Beneficiary	LEITAT
Nature of Deliverable	Report
Dissemination Level	Public (PU)

Submitted by	Harry Sarimveis, Irene Liampa, Periklis Tsiros, Philip Doganis, Pantelis Karatzas (NTUA), Dimitra Danai Varsou (NovaM), Nathan Bossa (LEITAT) and Iseult Lynch (UoB)
Revised by	Iseult Lynch and Anastasios Papadiamantis (UoB)
Approved by	Iseult Lynch (University of Birmingham)

Table of contents

Abbreviations

ADI - Allowable Daily Intake

AO - Adverse Outcome

AOP - Adverse Outcome Pathway

ATSDR - Agency for Toxic Substances and Disease Registry

BMD - BenchMark Dose

BMR - BenchMark Response

CI - Confidence Interval

CPD - Conditional Probability Distribution

CPT - Conditional Probability Table

DA - Defined Approach

DAG - Directed Acyclic Graph

DNEL - Derived No-Effect Level

ECHA - European Chemicals Agency

EFSA - European Food Safety Authority

EM - Expectation Maximisation

ENM - Engineered NanoMaterial

EPA - Environmental Protection Agency

FDA - Food and Drug Administration

FF - Far Field

FSLV - Final Safety Limit Value

GSD - Geometric Standard Deviation

IATA - Integrated Approaches to Testing and Assessment

ITS - Integrated Testing Strategies

JPD - Joint Probability Distribution

KE - Key Event

kNN – k Nearest Neighbours

LCIZ - Local Control Influencing Zone

LDA - Linear Discriminant Analysis

LLNA - Local Lymph Node Assay

LOAEL - Lowest-Observed-Adverse-Effect Level

LOEL - Lowest-Observed-Effect Level

MAD - Mutual Acceptance of Data

MIE - Molecular Initiating Event

MINLP - Mixed Integer NonLinear Programming

MLE - Maximum Likelihood Estimation

MOE - Margins of Exposure

MRL - Minimum Risk Level

MSDS - Material Safety Data Sheets

MSE - Mean Squared Error

MWCNT - Multi-Walled Carbon NanoTube

NF - Near Field

NM - NanoMaterial

NOAEL - No-Observed-Adverse-Effect Level

NOEL - No-Observed-Effect Level

OECD - Organisation for Economic Co-operation and Development

OEL - Occupational Exposure Limit

PBPK - Physiologically-Based Pharmacokinetics

PC - Phagocytizing Cell

PCA - Principal Component Analysis

PCF - Protein Corona Fingerprints

PEC - Predicted Environmental Concentration

PNEC - Predicted No-Effect Concentration

POD - Point of Departure

PSD - Particle Size Distribution

QSAR - Quantity Structure-Activity Relationship

RA - Risk Assessment

RAAF - Read Across Assessment Framework

RCR - Risk Characterization Ratio

RfD - Reference Dose

ROS - Reactive Oxygen Species

RP - Reference Point

SMAA - Stochastic Multicriteria Acceptability Analysis

TA - Transnational Access

UF - Uncertainty Factor

UI - User Interface

WBPBPK - Whole Body Physiologically-Based Pharmacokinetics

WoE - Weight-of-Evidence

Summary

The development of safe and effective nanomaterials (NMs) is highly important for both industry and regulatory agencies, especially considering their continuously growing economic potential, and their wide range of industrial, consumer, medical, and diagnostic NM applications. The basic methodology for performing risk assessment (RA) for NMs is similar to the philosophy used for conventional chemicals RA, i.e. compare the level of exposure with the hazard assessment. However, exposure and hazard assessments for NMs are more challenging than for conventional chemicals, because of the complex NM structures, which are dynamic as many of their properties are context-dependent (extrinsic), and can be modified or evolve during their life-cycle.

In this deliverable (D6.1) we describe a number of computationally oriented tools and methodologies that can be used for exposure modelling, hazard prediction and eventually for RA. Additionally, we present checklists and best practices for the most efficient use of the tools and workflows, as well as optimal combinations of these tools for performing RA for NMs. We report here on the current status of development and integration of existing RA tools into the NanoCommons knowledge infrastructure, and outline the strategy that will be used in the subsequent months of the project for further development, for supporting case studies to demonstrate the utility of the RA tools, and Transnational Access (TA) activities.

In particular, this deliverable presents the following modelling components that can either be used separately for particular tasks or in combination for the creation of complete RA workflows:

i)   The web-based GUIDENano guidance tool, part of the NanoCommons framework, which allows users to apply the most appropriate RA and risk mitigation strategy for NM-enabled products throughout their life cycle.

ii) Available tools and strategies for deriving points of departure (PODs) from dose-response data, i.e. levels of exposure that have low effect or no effect on humans or in the environment.

iii) The strategy for the development of novel grouping/read across approaches, which combined with the repository of nanoQSAR models described in deliverable report D5.4 offers a variety of NM hazard prediction tools.

iv) The tools that have been developed so far for the integration of biokinetics models and especially Physiologically-Based PharmacoKinetics (PBPK) models, which are excellent tools for estimating internal exposure of organisms to NMs.

v)   Bayesian modelling approaches, which can combine both hazard and exposure information and build complete RA workflows with the additional advantages that they can 1) provide predictions even in cases of missing data (which is typical for NMs) and 2) reveal information about mechanisms of actions. Bayesian models are perfectly combined with the emerging concept of Adverse Outcome Pathways (AOPs).

Integration of these models and RA tools early in the NanoCommons project lifetime facilitates their utilisation in the development of demonstration case studies to showcase the power and utility of the NanoCommons e-infrastructure for nanoinformatics to researchers, and more importantly, to industry and regulatory stakeholders, via the case study and stakeholder workshop activities of WP9.

1. Introduction

Nanotechnology is among the fastest growing technologies over the past few years, due to a wide range of applications of engineered NMs. However, the scientific community, the regulatory agencies and the industrial sector that designs and produces NMs are highly concerned about the potential adverse effects of NMs on human health and on the environment, which may be different to those arising from conventional chemicals or from micrometric particles (with bigger sizes). Due to their size in the nanoscale, NMs have a greater surface area, which enhances their chemical reactivity and may result in higher production of reactive oxygen species (ROS), and eventually lead to toxicity. The size and shape of NMs also allows them to move through the body and reach various organs and tissues. The distinct features of NMs and their effect on safety have attracted the interest of many researchers and practitioners in the nanotechnology area and drove the development of various RA frameworks, specific to NMs.

Oomen at al. [1] published a comprehensive and detailed review, addressing the aim, regulatory readiness, advantages, and disadvantages of 14 different RA frameworks and their applicability for NMs. All frameworks assessed followed the risk assessment paradigm, consisting of hazard identification, exposure assessment and risk characterisation, but differed in their aims, applicability domains, basic assumptions and alignment to one or more regulations. It is stated in this review paper that due to NM complexity it is not possible to construct an adequate RA framework to suit all routes of exposure for mammalian and ecological receptors.

In another review paper [2] the inclusion of *in silico* methods and approaches in RA frameworks is highlighted. Towards this direction, OECD has recently published a framework for Integrated Approaches to Testing and Assessment (IATA) [3], based on the concept of AOPs, which is based to a large extent on *in silico* modelling. The proposed framework combines nicely three different modelling methodologies to form a RA strategy. An exposure model predicts the external concentration, a toxicokinetics model determines the concentration of a substance in the various organs of the species of interest and the likelihood that a chemical reaches the target organs and a nanoQSAR or a grouping/read-across approach predicts if an AOP will be triggered.

Although the number of tools and models for NMs RA is growing, their use by industrial organisations and regulation agencies is not mainstream yet. Lack of sufficient data to support all steps required for RA is certainly one reason. We believe that one additional missing element is sufficiently described RA best practices and integrated user-friendly workflows that can guide users through the RA tools and provide navigational support on how to combine and link the different tools and approaches in order to arrive at reliable and well-validated RA and decisions. The NanoCommons project aims to develop pipelines and workflows that will fill this gap, and indeed to automate the whole process of RA, including provision of access to high quality data to run the models via the NanoCommons Knowledge Base.

At the heart of the NanoCommons RA infrastructure is the GUIDEnano tool, which is an extensive risk assessment tool, first developed during the EU FP7 project GUIDEnano. Since then, the GUIDEnano tool is continuously being extended and improved, through participation in various EU-funded projects. In addition, NanoCommons partners are developing and integrating several other modelling components that can be used either alone or in combination, to construct risk assessment pipelines. More specifically, NanoCommons partners are designing novel grouping/read-across approaches based on

mathematical programming, which automate and optimise the Read Across Assessment Framework (RAAF) proposed by ECHA and has been particularly tailored to the specific requirements of NMs [4]. Partners are also developing tools for creating and integrating biokinetics models in the Jaqpot modelling platform, which is also a major part of the NanoCommons e-infrastructure. Through Jaqpot, we are also integrating the PROAST software developed by RIVM is also being integrated into the NanoCommons RA workflow. PROAST offers an implementation of the BenchMark Dose (BMD) approach, which is the most preferred dose-response assessment method for calculating Derived No-Effect Levels (DNELs) based on dose-response data. Work has also started on developing services for creating and hosting Bayesian causal models, which is an alternative approach to construct RA frameworks combining into the same network exposure and hazard information and RA.

As highlighted above, in this deliverable the aforementioned tools are described, their current status of implementation and integration documented, and workflows that can combine these tools in the most efficient way in order to arrive at reliable and validated decisions are presented. The task and WP6 partners are also in the process of describing the next development steps (to be implemented in the second period of the NanoCommons project) and how the different tools can serve the TA needs and WP9 requirements, via collaboration with WP9 partners to develop and implement demonstration case studies for NMs RA informatics approaches.

2. General Workflow

General concept

NanoCommons will provide a unified, structured, end-to-end approach for NMs RA by linking together strategies for exposure assessment, hazard identification and quantification and risk analysis. The goal of the NanoCommons e-infrastructure platform is to offer a variety of nanoinformatics-based solutions for RA depending on the desired level of detail and the availability of relevant reference information (in-house from the user, or accessed via the NanoCommons Knowledge Base, see also Deliverable Report D4.4 on the first implementation of the NanoCommons Knowledge Base). The NanoCommons approach integrates state-of-the-art tools and methodologies in a hierarchical way as a means to perform the risk calculation in an evidence-based manner, following the standard REACH and ECHA guidance frameworks.

Course of action

NanoCommons partners opted to focus on quantitative RA throughout the NM life cycle using a process-specific basis, calculating initial exposure estimates, hazard reference points and the respective Risk Characterization Ratios (RCR), yet allowing for the assessment of the effect of a range of mitigation measures (Figure 1) to reduce the overall final risk. The various methodologies involved in Figure 1 are described next. More detailed information on the NanoCommons tools that are used to implement the various stages in this workflow, such as GUIDENano and predictive models hosted on Jaqpot platform, are reported in Sections 3-7 of this deliverable.
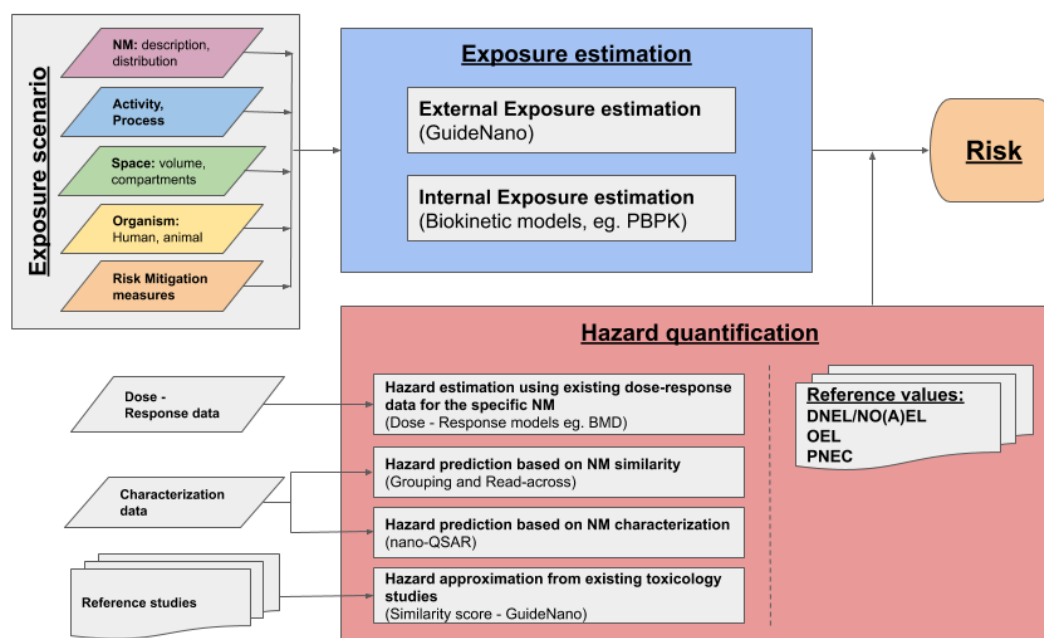

Figure 1: Overview of the NanoCommons nanoinformatics based RA framework. A key aspect of the workflow is integrating the different aspects, and automating the process to enable the RA to be run end-to-end.

Exposure estimation

At the core of this approach is GUIDENano, a web-based guidance tool developed through the EU FP7 project GUIDEnano, that performs in-depth RA of NMs and nano-enabled products throughout their life-cycle, i.e. from production to waste management. For this purpose, it requires basic information about the NM, such as the physicochemical properties, composition, size and quantity, to be provided as input by the user. In order to determine the exposure and hazard, the GUIDEnano tool offers an activity list (i.e. all processes involved during the NMs life cycle such as pouring, abrasion,...) spanning more than 200 processes to choose from, and includes a collection of state-of-the-art exposure models that take into account the NM's fate and the changes that NMs undergo in different environments. The GUIDEnano tool currently performs very detailed external exposure simulations that enable prediction of environmental concentrations (Predicted Environmental Concentration, PEC) or human exposure (for example ingested, dermally applied or inhaled dose of NMs following the release of NMs in a working facility or in the environment).

When it comes to human risk assessment, the internal exposure is the main driver of the risk. Therefore, for the most accurate determination of the health risks of NMs and to perform a robust human RA, it is essential to estimate the internal exposure [5]. Internal exposure is the mass or concentration of the NMs that reaches the systemic circulation, the organs and tissues of an individual, following the injection, ingestion or inhalation of the NM. In order to perform internal exposure estimations, information on the toxicokinetics of the NMs (the absorption, distribution, metabolism and excretion processes) is needed. Excellent modelling tools for such simulations already exist, including compartmental models and more specifically the Physiologically-Based Pharmacokinetic (PBPK) models. PBPK models allow RA to be performed on an individual basis, by adapting the physiological parameters to the personal characteristics of each individual (sex, age, weight). They can also fully integrated and combined with the concept of AOPs, as they provide concentration-time profiles for all important organs of the individual and thus for the organs where AOPs are initiated. Within the NanoCommons project, and in particular through the Jaqpot computational platform, we infrastructure for developing, storing and sharing PBPK models has been developed, which is being integrated into the overall RA framework, as illustrated schematically in Figure 1.

Hazard assessment

Hazard assessment is concerned with the determination of threshold limits that define safe use and exposure levels for NMs. According to REACH, the following occupational and environmental threshold limits (Final Safety Limit Values (FSLVs)) represent the reference values for assessing whether risks from NMs (or indeed other chemicals) are controlled:

- Predicted no effect Concentration (PNEC): Concentration of the substance below which adverse effects in the environmental sphere of concern are not expected to occur; PNECs can be derived for long term or short term exposure).
- Derived No-Effect Level (DNEL): Level of exposure to a substance above which humans should not be exposed; DNELs need to be derived for each relevant exposure pattern (population, route, and duration of exposure) and each relevant health effect (local and systemic effects).
- Occupational Exposure Limit (OEL): maximum admissible concentration at workplace, averaged over a specific period of time.

In some cases, these limits are available in material safety data sheets (MSDS), and/or the registration dossiers for REACH. However, in most cases, due to the absence of formal limits, and the lack of available epidemiological data, the FSLV should be estimated from any available information. In NanoCommons three alternative but complementary methodologies for hazard quantification are considered and integrated:

1. Dose response modelling: On the condition that toxicological dose-response data for the NMs are available, Dose-response modelling and specifically the BenchMark Dose (BMD) approach can provide the dose corresponding to a predefined BenchMark Response (BMR), which is the concentration (typically a percentage, eg. 5%) where adverse effects start to emerge, evaluating the parameters that contribute to it at the same time. The BMD is derived from the estimated dose-response curve, taking into account the respective statistical uncertainty in the estimate and considering all models that are compatible with the data. In practise, the lower confidence limit, namely the BMDL, is provided to the user to be further employed as a reference point (RP, also called the point-of-departure (PoD)), to derive a health-based guidance value. For this purpose, we have integrated the gold standard PROAST tool, available from RIVM, on the Jaqpot platform, offering a complete BMD workflow.'

2. Predictive modelling: There are two methodologies for developing models for predicting adverse effects of NMs: nano-Quantitative-Structure-Activity-Relationship (nanoQSAR) models and read-across methods. The nanoQSAR modelling method is discussed and the relevant tools offered by NanoCommons partners were described in detail in deliverable report D.5.4. Read-across non-testing strategies are employed for the prediction of NM toxicity, in cases where sufficiently large datasets are not available for the development of reliable nanoQSAR models. This approach is grounded on the empirical knowledge that similar materials may exhibit comparable properties and thus, the estimation of the hazardous effects of non-tested NMs can be achieved using data within a group of comparable NMs (ECHA, refers to similar NMs of the same composition as nanoforms, and currently only supports read-across within a single NM core composition). There are two main read-across approaches; the grouping and the analogue approach. In the grouping approach, the NM samples are organized into groups of similar compounds. Groups are formed considering structural similarities between samples, and it is assumed that due to these similarities, the biological or toxic activity of the NMs within a group follows a regular pattern. In the analogue approach the prediction is limited to a small area of the data space; one source NM can be used for the endpoint estimation for a single or more target NMs, or two or more source NMs can be used to make predictions for a single or several target NMs. In this approach, the endpoint prediction can be achieved either by applying QSAR methods locally or by implementing a (k nearest neighbours) kNN-like algorithm to experimental observations of only a few neighbours to the query NM, in order to compute the endpoint prediction. In NanoCommons, we are currently developing a novel methodology based on mathematical programming, which keeps only the most informative NM descriptors coupled with the optimal definition of the "neighbourhood" around the NMs that are considered to be similar to the target NM. We are also offering a read-across methodology which is based on kNN clustering. Additional methodologies are being developed by project partners and will be presented in detail in the subsequent deliverable report D6.2.

3. Similarity assessment rules: Another strategy for hazard assessment is the approximation of FSLVs by employing a set of similarity assessment rules for human and environmental hazard from

relevant literature. The GUIDENano platform offers a sophisticated workflow for the assessment of the quality, relevance and similarity between the NMs for the candidate toxicity studies that are to be used as FSLV resources. The result is a ranked list of the acceptable resources and the respective reference values. The corresponding tool submodule also provides further refinement of the surrogate FSLVs by taking into account the uncertainty introduced by the source of the respective information. This is achieved through the quantification of the dissimilarity between the studies and the exposure relevant material by deriving an uncertainty factor, used subsequently for deriving a "corrected" FSLV.

Risk assessment

Finally, a risk score, i.e. a risk characterization ratio (RCR), is calculated for each exposure scenario, by comparing the measured or estimated exposure levels and the PNECs for the environment and DNELs for human health [6], according to Equation 1:

$$RCR\ human/occupational\ = \frac{exposure}{DNEL\ or\ OEL} \quad (1)$$
$$RCR\ environmental\ = \frac{PEC}{PNEC}$$

The risk to humans can be considered to be controlled/acceptable if the estimated exposure levels do not exceed the appropriate DNEL (i.e, exposure estimate/DNEL<1). Similarly, the risk to the environment can be considered to be acceptable if the PEC values do not exceed the PNEC (i.e, PEC/PNEC<1).

The GUIDENano tools offers many options for such calculations. For example, to assess a worker's exposure, GUIDENano calculates different short- and long-term exposure estimates and RCRs for the different exposure routes, eg. inhalation. The user can also choose from a range of exposure modifiers that establish the realistic operational conditions and assess their effect, in the final exposure scenario, using the GUIDENano tool. For environmental exposure assessment, a separate risk score for each environmental compartment is calculated. When the risk analysis is completed, a highly informative graphic is generated by GUIDENano where the risk scale is represented by a chromatic band where the green colour corresponds to an RCR value <1, that means that no risk is present under the selected conditions whereas the red colour indicates an RCR value >1 that corresponds to the presence of risk for human health or the environment. In the latter case, guidance on risk mitigation measures are provided based on the specific NM, product and exposure scenario.

Workflow description

The complete workflow for RA, from the user point of view, is depicted in Figure 2. The workflow guides the user throughout the process of collecting the relevant information efficiently, allowing for the creation of multiple elaborate exposure scenarios (see Appendix 2 - Checklist R1-R8). Data on the NM physicochemical properties, exposure conditions, toxicokinetics, fate and hazard related to the given NM(s) are incorporated into each scenario.

If both the exposure level and the hazard reference point are available then the RCR is calculated in one simple step. Apart from the measured exposure level, the user may themselves define an exposure value in a way that various hypothetical scenarios can be also evaluated (see Appendix 2 - Checklist R9, R12).

However, since the availability of the required information is usually limited, in most cases the user needs to calculate an estimate for the exposure level and thus is prompted to enter the collected information into the GUIDENano platform. A collection of sophisticated exposure models, hosted in GUIDENano, produces accurate estimates of the external exposure level, taking into account the given exposure conditions (see Appendix 2 -Checklist R7). These outputs include the NM release activity, the NM life cycle, the relevance to the NM source, the exposure route and any exposure modifying parameters (see Appendix 2 - Checklist R2-R6 for further details).

In case the internal exposure levels to specific organs or systems needs to be estimated, the relevant information is used to run the appropriate PBPK model, integrated in the Jaqpot platform, depending on the exposed organism, the exposure route and the organ/system of interest (see Appendix 2 - Checklist R8).

To overcome the limitation of low availability of hazard RPs, the NanoCommons workflow integrates multiple hazard quantification information sources in a hierarchical manner depending on their relevance, quality and accuracy. Thus, the user can choose between the BMD approach (Appendix 2 - Checklist R10) offered through Jaqpot, the Read-across models (Appendix 2 - Checklist R11.2-R11.3) also hosted on the Jaqpot server, or/and the GUIDENano similarity module (Appendix 2 - Checklist R11.4). The comprehensive description of the aforementioned methodologies, models and tools is encompassed in the following sections of this deliverable report.

The final output is a list of FLSVs according to the confidence level corresponding to each particular resource. The user can choose to use any value available in the list for the RCR computation, either the minimum one, representing the worst-case scenario, or the most relevant to the purpose of the RA (Appendix 2 - Checklist R13). Eventually, the user may select available protective controls or exposure modification factors to evaluate their effect on the RCR for each assessed exposure scenario (Appendix 2 - Checklist R14).
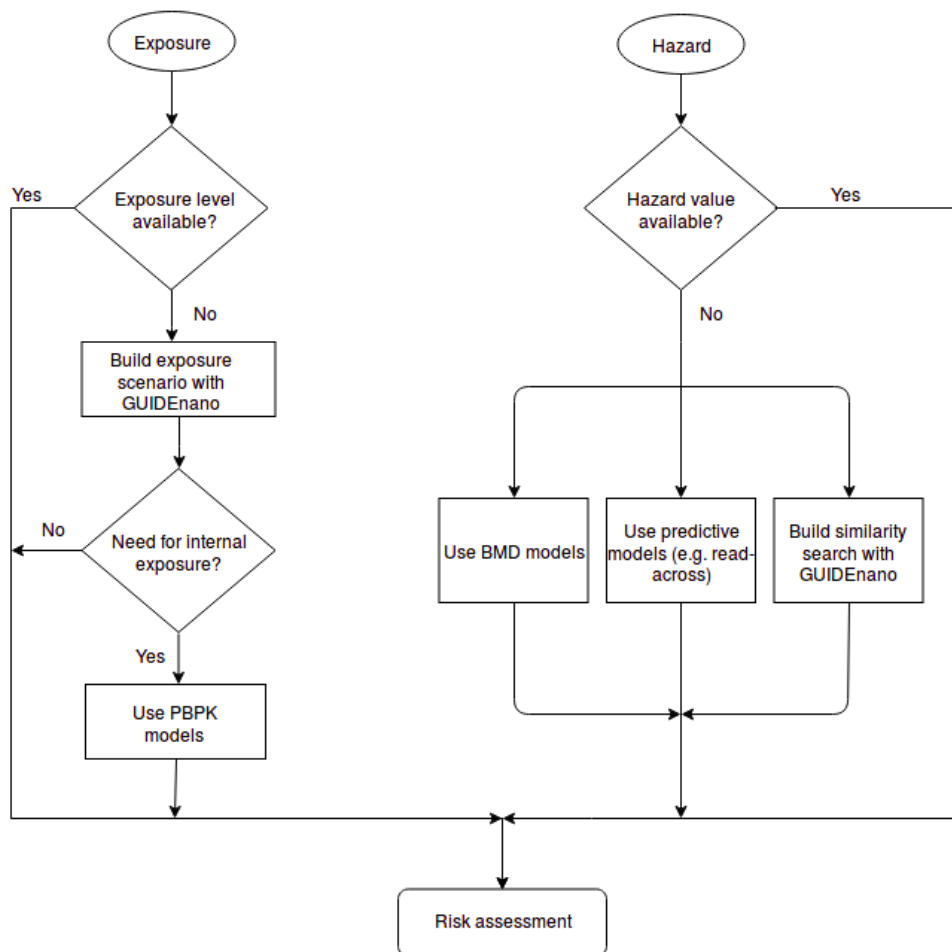
Figure 2: Flowchart of the NanoCommons nanoinformatics-based RA workflow.

3. GUIDEnano tools

GUIDEnano is an interactive web-based Guidance Tool that aims to guide NM producers and NM-enabled product manufacturers towards the safe and sustainable design of their NM-enabled products [7]. GUIDEnano combines a range of predictive models, multi-level decision trees, and databases to derive critical information along the RA and risk mitigation processes. In this deliverable we provide an overview of the GUIDEnano conceptual framework and the different modules of the GUIDEnano Tool (see Figure 3). The GUIDEnano Tool is still under development, the most relevant ongoing activities are connection with existing databases (NanoCommon), the generation of new tailored databases to cover existing information gaps (NanoCommons and caLIBRAte EU projects), and additional developments to facilitate the usability of the Tool and the incorporation of new research findings to refine the risk assessment process (Gracious EU project).
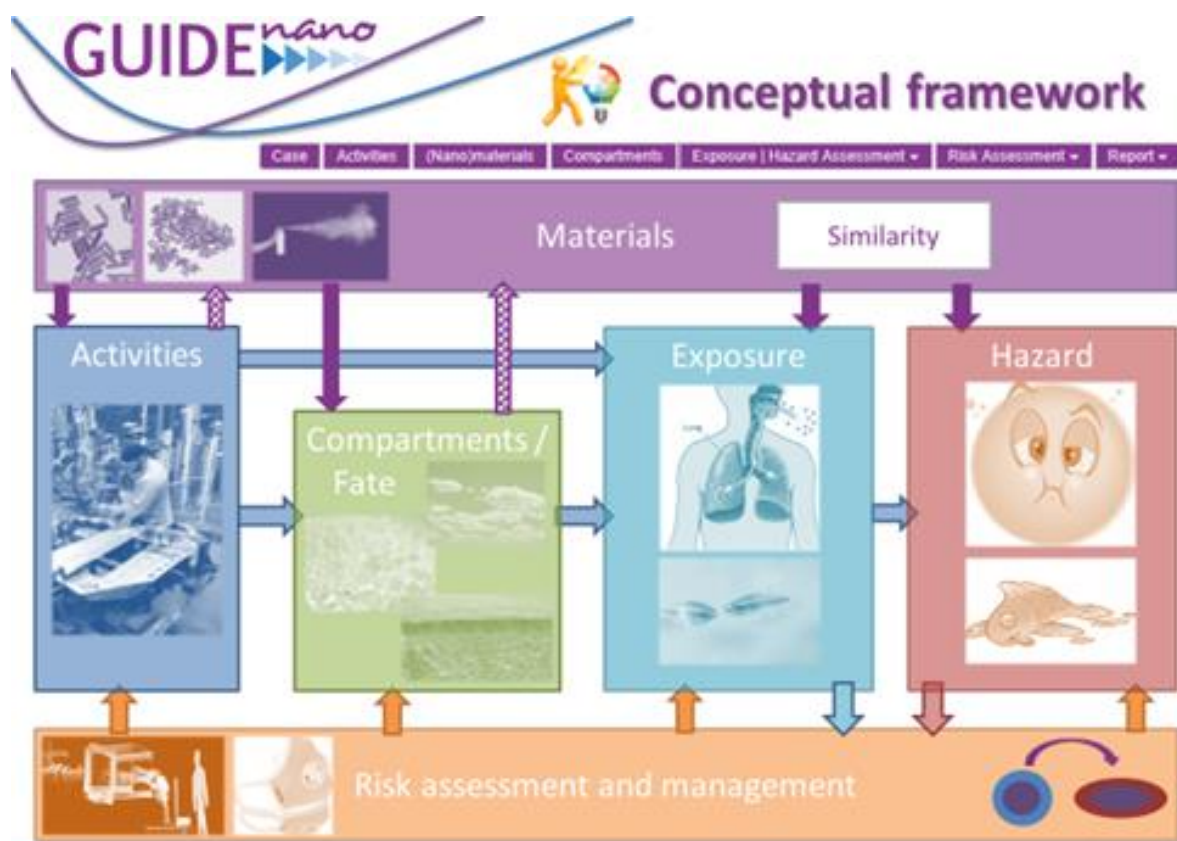

Figure 3: General overview of GUIDEnano conceptual framework.

Application module

The application module is server based and allows multiple users to access the application and work together on a single case in real time. After login, the user can create a case study, work on previous case studies, and share the case study with other users.

Life cycle activity module

The life cycle activity module allows a user to define all relevant activities within all stages of the life cycle of the NM and nano enabled product/article[1]. An activity can be defined as every process/action occurring during the NMs life cycle, as shown schematically in Figure 4.

Figure 4: Exposure along the NMs life cycle, where each step / process is considered to be an activity that may generate a release of / exposure to the NM in some form.

The scope of the RA may vary from a single activity up to all activities within the entire product life cycle. An 'activity' within the tool is an intentional, man induced and controlled process, such as: 'mixing', 'drilling', 'grinding', 'knitting', 'wearing', 'washing', 'spraying', 'painting' etc. An activity is not necessarily associated with a release of NMs but can be just a description of an action, for example, transport of paint containing NMs from the factory to the users' house. Also, another vision is that the activity can describe a modification of the state/form of the NMs during a process. For example, an activity can describe the pouring of nano-TiO_2 into a paint matrix and may describe the associated release of nano-TiO_2 to the air as well as indicate that the NM is now in another state as a substance/mixture[2]. An important output from the activity in terms of RA is that the activity describes the NM mass balance with an input, an output and a (potential) release, as shown in Figure 5.

---

[1] An article under the EU's REACH regulation is defined as '…an object which during production is given a special shape, surface or design which determines its function to a greater degree than does its chemical composition' (Article 3(3)).

[2] A substance under REACH means 'a chemical element and its compounds in the natural state or obtained by any manufacturing process, including any additive necessary to preserve its stability and any impurity deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition' while a mixture in this case is an intentionally prepared blending of two or more constituents (also called a 'preparation'). See also: http://www.refac.eu/usefulinformation/definitions.aspx

Figure 5: Example of activity input parameters for TiO₂ NMs synthesis and results showing the calculated release into compartments (workplace) and consequent exposure to the worker.

It should be noted that an activity is described as a function of its scale, the energy involved, life cycle stage, several parameters to characterize its duration and frequency, redundancy, and materials release rate. The number of activities is not limited and can describe complex processes such as weathering of an outdoor painted wall during a use time of 20 years as well as simple process like paint drying. The Activity framework is directly correlated to the (nano)materials and compartment frameworks, described below.

(Nano)material module

The (nano)material module allows a user to define all relevant materials: substances, nano-objects, mixtures, and nano enabled products (or articles) present during the NMs life cycle. User defined materials can be structured/mixed and reused as a constituent of another material using the 'proxy' concept. All defined materials need to be described including the under-lying components, chemicals and their role. Physical, chemical and toxicological parameters can be included for each material (e.g. shape and size, physical and surface properties, functions, chemical info, reactivity info, classification and labelling, toxicity studies). The user can add characterization data for any input (nano)material involved in the activity, for example the nano-TiO_2 incorporated into paint and the paint matrix, but also characterization data for the materials generated during an activity, for example, the release of nano-TiO_2 embedded in residual weathered paint matrix. The parameters used to describe the materials are different if it is an article, substance, mixture, nano-object, nano-structured aggregate or nano-structured agglomerate. On the basis of the material constituents, information entered by the user, the tool generates a series of graphical schemes to illustrate the NM composition and arrangement (e.g. core, shell, wall or coating) including a 2D schematic view of the described materials (Figure 6). GUIDEnano also provides a quality criterion (0 to 100%) that informs the user on the quality of data entry for each material described. The quality score reflects only the level of characterization of the materials and is only informative for the user. A good quality score is not a requirement to pursue the RA. The materials defined are used in the other modules.
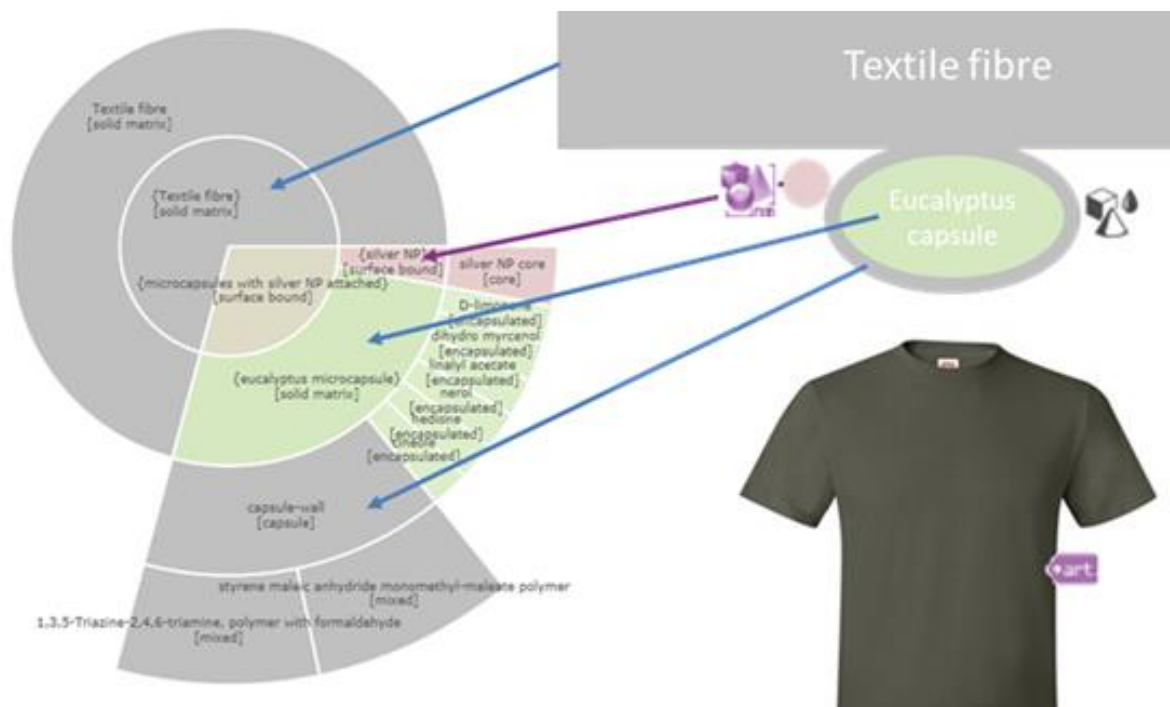
Figure 6: Example of 2D representation of materials.

Compartment module

The compartment module allows the user to define all compartments and zones involved during the NMs life cycle, from the factory, toward the house or the landfill. There are two groups of 'compartment-classes' to consider: 'system' and 'environmental'. The 'system' compartment-classes are manmade like: 'sewage system', 'wastewater treatment plant', 'indoor air (room)' and 'landfill site'. The group of 'environmental' compartment-classes currently in the tool are: 'fresh water', 'estuarine', 'marine', 'fresh water sediment', 'salt water sediment', 'outdoor air' and 'soil'. The user can create as many compartments as relevant for the scope of their case study. To help the users, the compartments are already prefilled with standard values (size, composition, properties), however the user can modify the compartment characteristics to fit is own simulation. The user can also specify some 'zone' which is a logic area/region within a 'compartment' (Figure 7).

Figure 7: Factory compartment and zone description.

The user can describe the position/size of the zone in the compartments and create connections between the zones and between the compartments to assess any exchange. in Figure 7, the compartment factory hall is described with the zone of the floor, the near field zone (NF), and the far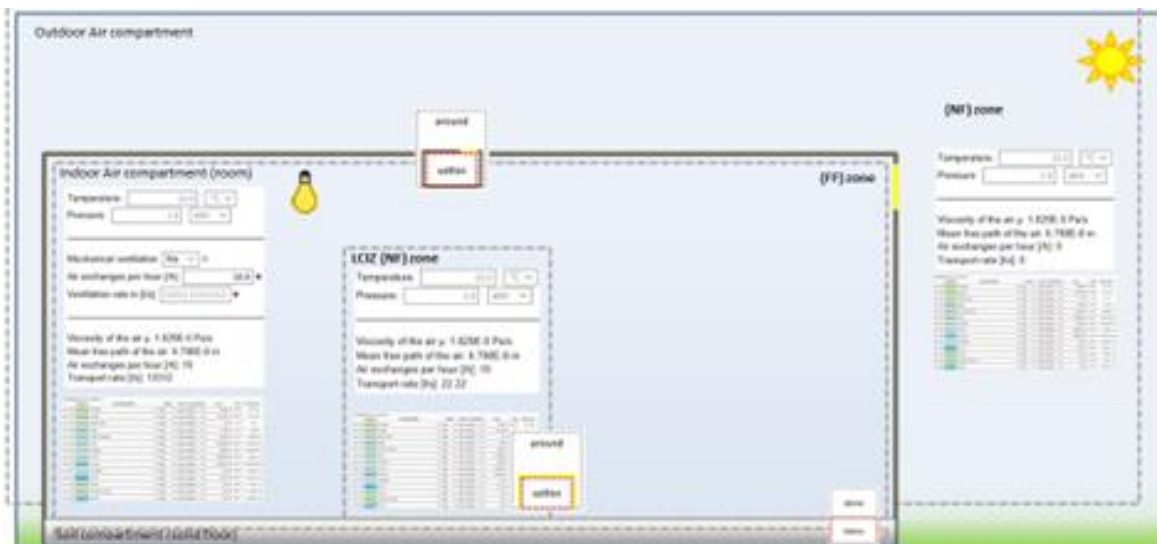 field zone (FF zone, representing the remaining of the factory room) and outside the factory hall. In addition, in the Two-Box Nano-specific model developed to model the aerosol behaviour (exposure module), we located a box around the source and call this the local control influencing zone (LCIZ) to differentiate from the workers-NF zone and the remainder of the room (FF).

Exposure and Hazard module

From the information input into the activity, materials and compartment module, GUIDEnano determines the predicted environmental concentration (PEC) in all compartments and zones described. The exposure module enables the user to define multiple 'exposure scenarios' for both 'human populations' and 'eco-populations' throughout the entire product life cycle. For each compartment the tool automatically introduces a new 'eco-population' for it. A PEC is derived for all defined zones in each compartment (Figure 8). To obtain the PEC, the user can use the internal model implemented in GUIDEnano or enter data obtained from different methods or experimental data. For occupational settings, GUIDEnano implemented the Two-Box Nano-specific model (see further description below) to estimate the evolution of particle concentration in air with time. The choice of the models to estimate the PEC can be different for each exposure path.

Two-Box Nano-specific model

The model is based on the Near-Field (NF) Far-Field (FF) source receptor model developed by Cherrie [8, 9] and the algorithms developed by Maynard and Zimmer [10] for estimating the time evolution of the particle size distribution (PSD) of NMs. In a two-box (or source-receptor) model the room is typically split into two boxes, one is placed either around the worker or NF source, while the other is the remainder of the room (FF). We located a box around the source and call this the local control influencing zone (LCIZ) to differentiate from the workers-NF zone and the remainder of the room (FF), which is referred to as in the original model by Cherrie (Figure 8).
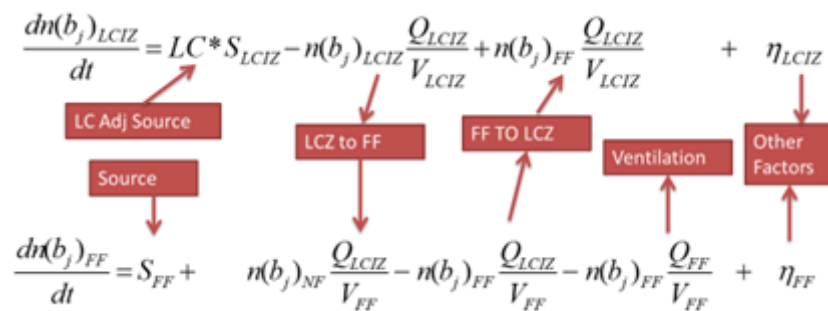
Since the worker moves around the room, the concentration in the worker NF zone can be estimated from the concentrations in the LCIZ and FF zones and the time the worker spends in each zone. The concentration in either zone is then a function of agglomeration, diffusion, dispersion, dilution and emission, with each size bin in each zone (LCIZ or FF) being described using the following equations:

$$\frac{dn(b_j)_{LCIZ}}{dt} = LC * S_{LCIZ} - n(b_j)_{LCIZ}\frac{Q_{LCIZ}}{V_{LCIZ}} + n(b_j)_{FF}\frac{Q_{LCIZ}}{V_{LCIZ}} + \eta_{LCIZ}$$

LC Adj Source

Source

LCZ to FF

FF TO LCZ

Ventilation

Other Factors

$$\frac{dn(b_j)_{FF}}{dt} = S_{FF} + n(b_j)_{NF}\frac{Q_{LCIZ}}{V_{FF}} - n(b_j)_{FF}\frac{Q_{LCIZ}}{V_{FF}} - n(b_j)_{FF}\frac{Q_{FF}}{V_{FF}} + \eta_{FF}$$

where LCIZ is a virtual zone that represents the area of influence of any exposure control around the source, Q_{LCIZ} is the volume air flow between the LCIZ and FF (m³/sec), Q_{FF} represents the ventilation (m³/sec), and represent the mass emission rate into the LCIZ and FF, respectively in particles/sec, LC represents the local control adjustment factor and η represents the size-specific factors (coagulation, diffusion and dispersion).

$$\eta = \frac{dn(L_j)}{dt} = \frac{1}{2}\int_{i=1}^{j} n(L_i)\,\beta(L_i, L_j - L_i)\,n(L_j - L_i)\,dL_i - n(L_j)\int_{i=1}^{N} \beta(L_j, L_i)\,n(L_i)\,dL_i$$

$$- \frac{n(L_j)v_{ts}}{h} - \frac{A\,n(L_j)\,D}{V\,\delta_{diff}}$$
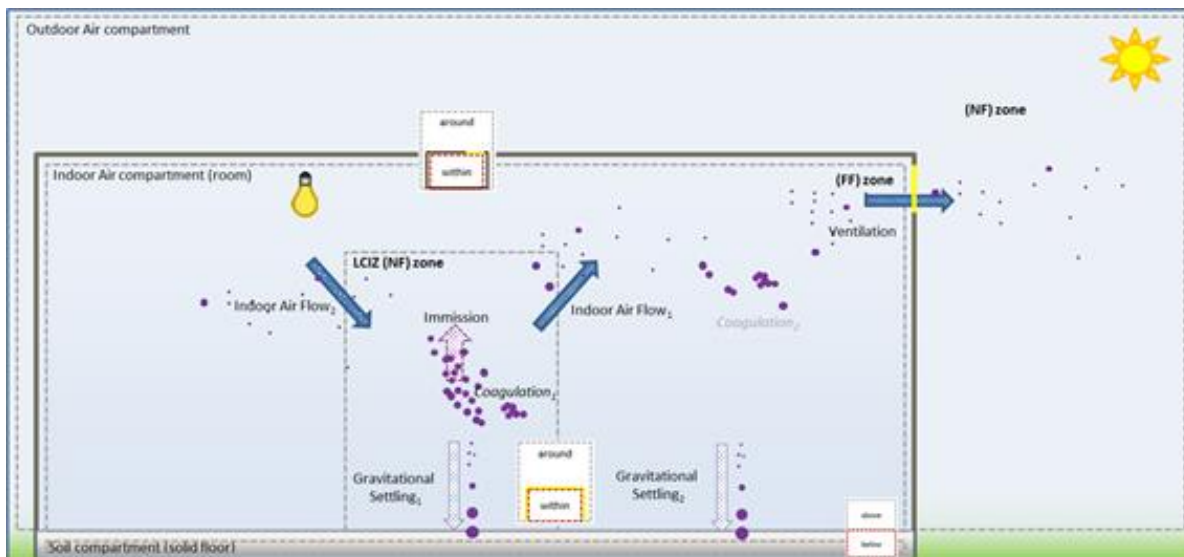
Figure 8: Exposure of NMs in the factory compartment using the two-box nano-specific model.

For human exposure, the GUIDEnano exposure scenario is characterized by the type of exposure (i.e. indirect or direct), the route of exposure (i.e. inhalation, dermal or oral) as well as the materials involved (see materials module). For each defined exposure path, the tools estimate(s) the PEC, which is calculated over time and the tools calculate the concentration peak (maximum) and the long-term concentration by averaging values over all timepoints. The user can include personal protective equipment (mask, gloves, etc.) within the exposure module and/or general exposure controls (e.g. local exhaust ventilation) within the activity module and quantify their effect on exposure. Details about the connection and between Materials, activity, compartment and exposure modules are presented in Figure 9.
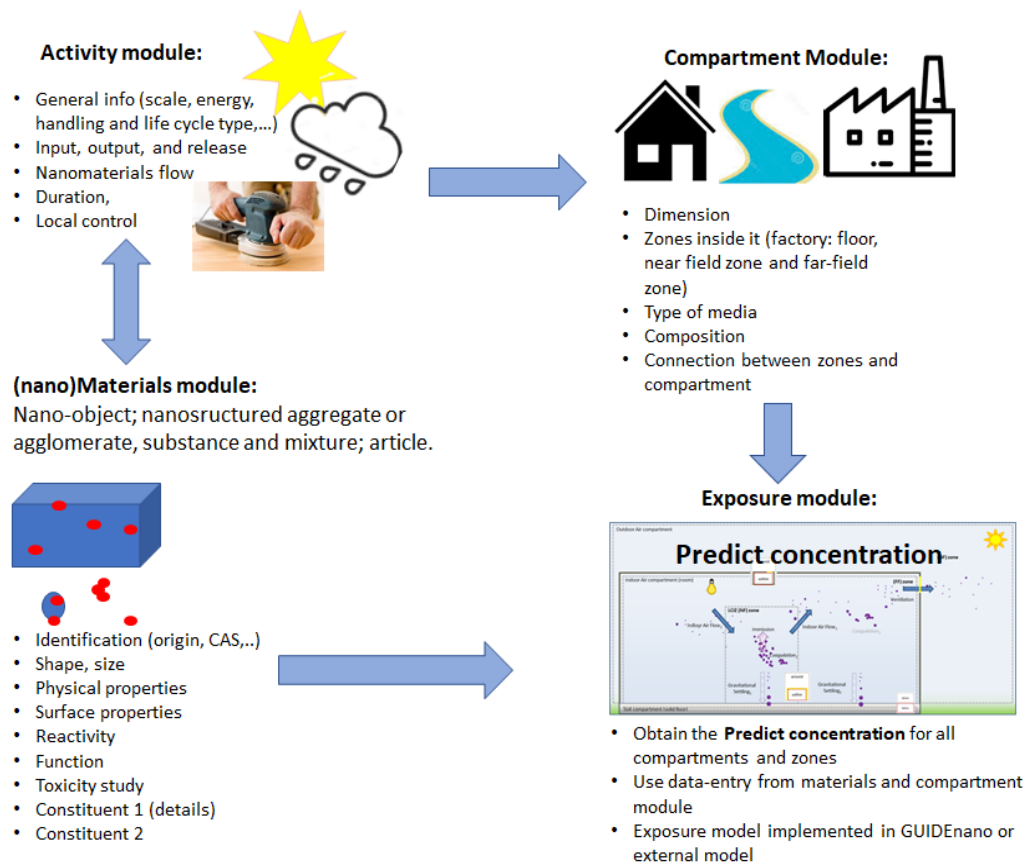
Figure 9: Workflow to obtain the predicted concentration value from the exposure module.

Hazard assessment is presented as a tiered approach to obtain a Final safety limit value (FSLV) (Figure 10). In its current version, Step 2 and Step 3 are interchanged, and step 3 in Figure 10 is presented as Step 2. This will be modified in the future, to align the system to this figure and to improve usability of the tool. The description below refers to the currently implemented structure for hazard assessment.
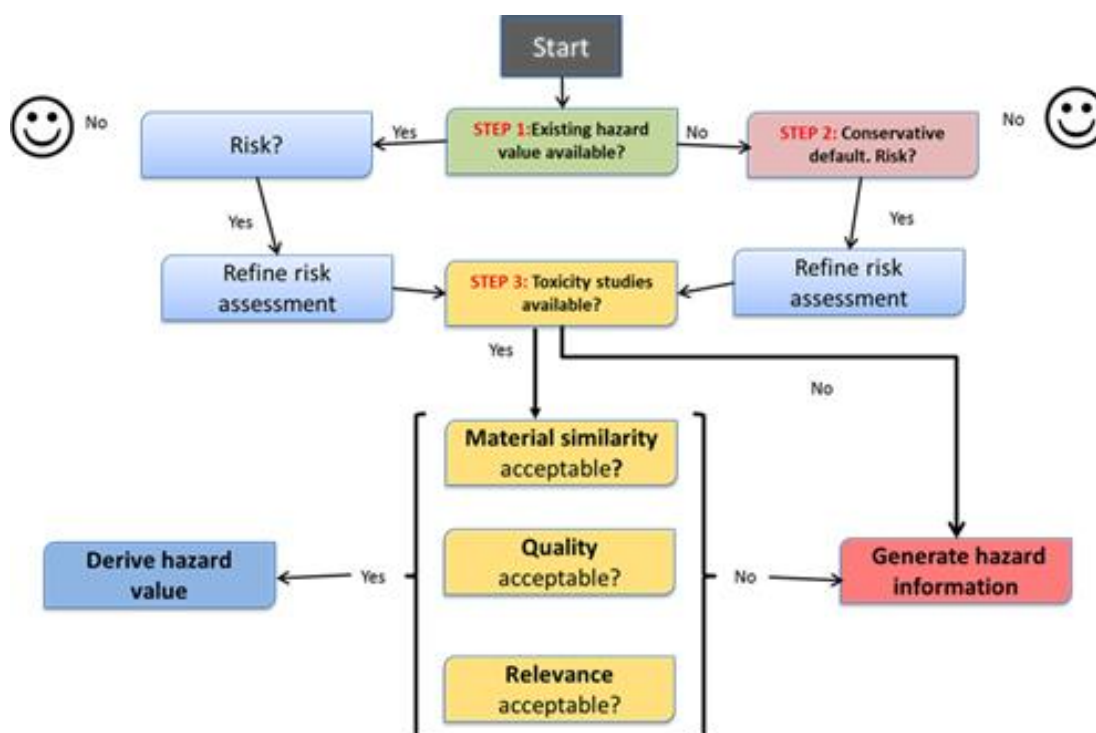
Figure 10: Decision tree developed in GUIDEnano to obtain a FSLV value as the output from the RA.

In STEP 1, the user is asked if an existing hazard threshold value (i.e., DNEL or OEL-like, for example from a regulatory body or research publication) for the exposure relevant material already exists. If so, the user can add this existing safety limit value to be used as the FSLV, otherwise STEP 2 is activated. In the case the user used a worst case DNEL value in step 1 and 2 and a risk is considered, the user can refine the DNEL value to a more relevant one by going to STEP 3.

In STEP 2: The user is asked if there exist relevant 'toxicity studies' for the 'exposure relevant' or a 'similar' material. If this is the case, the user is offered a list of relevant toxicity study templates to choose from to enter the toxicity study information and results. Relevance, quality and similarity are used to conclude on the acceptability of a toxicological study for a given hazard endpoint. Only 'accepted' studies are used to derive the FSLV. By similarity, the tool refers to the similarity of the exposure relevant material versus the material used in the toxicology studies [11, 12]. In the similarity tool, the 'context' determines which material characteristics to compare. For example, two particles may behave the same as aerosol but completely different once dispersed in water. Therefore, a list of different contexts is proposed. For instance, "inhalation toxicity", "aquatic fate" etc. An 'index' is a matching condition related to a single or group of related properties. Think of "melting point", "shape and aspect ratio", "size distribution, size type and method". Other contexts include shape, crystallinity, size distribution, physical state, chemical composition. Each 'index' implements its own 'matches' function. This 'matches' function contains the expert rules/algorithm on how to compare two materials for that specific 'index', regardless of the context. Comparing the same 'index' of two materials must lead to a context independent score between 0 and 1. The final similarity score will be derived from the combined and weighted index scores (see details below). In the RAt module this tool will be used in case no toxicity data is available in the literature for the same NMs. When several studies are accepted for a hazard endpoint, the study with the lowest DNEL is selected as the FSLV.

Similarity score

The final similarity score will be derived from the combined and weighted index scores. In the current version, the weights are all 1 and the final score is equal to the lowest scoring condition.

Constituent matching

A complicating factor are the differences in material compositions and structures to compare. If two materials have a different number of constituents and concentrations, matching the constituents to compare is a real challenge. Three constituent properties: 'concentration', 'role' (of the constituents) and 'surface ordering' might be used to determine which constituents to compare.

Identifying chemical compound(s)

The first step is determining the 'identifying chemical compound(s)' of a material. This might seem trivial as people often talk about a material implicitly referring to just one of its constituents. For example, we talk about silver NMs because the silver gives the NM it's specific function and not the 'coating' or impurities. This identification process needs to be formalized into a set of explicit rules. The first rules implemented for now are:

The identifying chemical compound(s) of a primary nano-object are those chemical constituents with their roles: 'core' or 'wall'. In the case of 'agglomerates' or 'aggregates', the identifying chemical compound(s) are derived from the aggregated/agglomerated particles.

Comparing 'chemical composition'

The chemical composition of two materials (m_1, m_2) is compared by comparing the 'identifying chemical compound(s)' of both materials. The current version only supports comparison of two materials with one 'identifying chemical compound' each. We assume $chem_1$ and $chem_2$ to be the 'identifying chemical compound(s)' to compare. The first step is comparing CAS number(s), EC number(s) and/or chemical formulas. If they match the difference in their respective 'mass percentages' is used to calculate the score. Otherwise the score = 0.

Comparing 'physical state'

Five different 'physical states' are defined within the tool: 'ions', 'solid', 'liquid', 'gas' and 'mixture'. For now, comparing the 'physical state' of two different materials is implemented in a straightforward manner whereby only materials in the same state are considered similar at the moment. Also, the ionic form is currently considered equal to a 'solid', 'liquid' or 'gas' state.

Comparing 'crystallinity'

In the case of two solid materials the structure of both materials is matched. If both are (poly)crystalline each of the 'crystallinity forms' and 'space group' are compared based on their percentages.

Comparing 'Size distributions'

A 'size distribution' similarity algorithm has been developed based on comparing the cumulative size distributions. The difference between the two size distributions is related to the surface area between the two cumulative graphs. A bigger surface area indicates a lower similarity, and vice versa.

If none of the available studies is acceptable STEP 3 is activated, which is based on worst case default hazard thresholds for generic groups of NMs. For STEP 4 the quantitative 'final safety limit value' or the qualitative 'outcome' is passed to the risk assessment module.

Risk assessment module

The risk assessment module combines the exposure and hazard module entry and output information, providing a RA for all defined materials in each compartment as shown in Figure 11.

Figure 11: Tiered approach used in GUIDEnano to obtain the most relevant FSLV value.

In step 4, the tool generates a matrix overview of all assessed hazard endpoints for all different exposed populations in the life cycle considered. For each endpoint used for the hazard assessment, GUIDEnano combine the exposure and hazard data to obtain an RA analysis. A red balance indicates that the probability of risk is considered "high", an orange one "medium", a green one "low" and black indicates "not determined yet" (Figure 11). By clicking the balance-icon you can open a dialog with a more detailed RA description (Figure 12):

Figure 12. Overview of all assessed hazard endpoints for all different exposed populations involved in the simulation.

As part of the 'risk management strategy' the tool can be used to see the effects of certain strategies of risk mitigation on the risk outcome. As an example, the effect of wearing gloves on the risk is shown in Figure 13.

Without gloves   With gloves
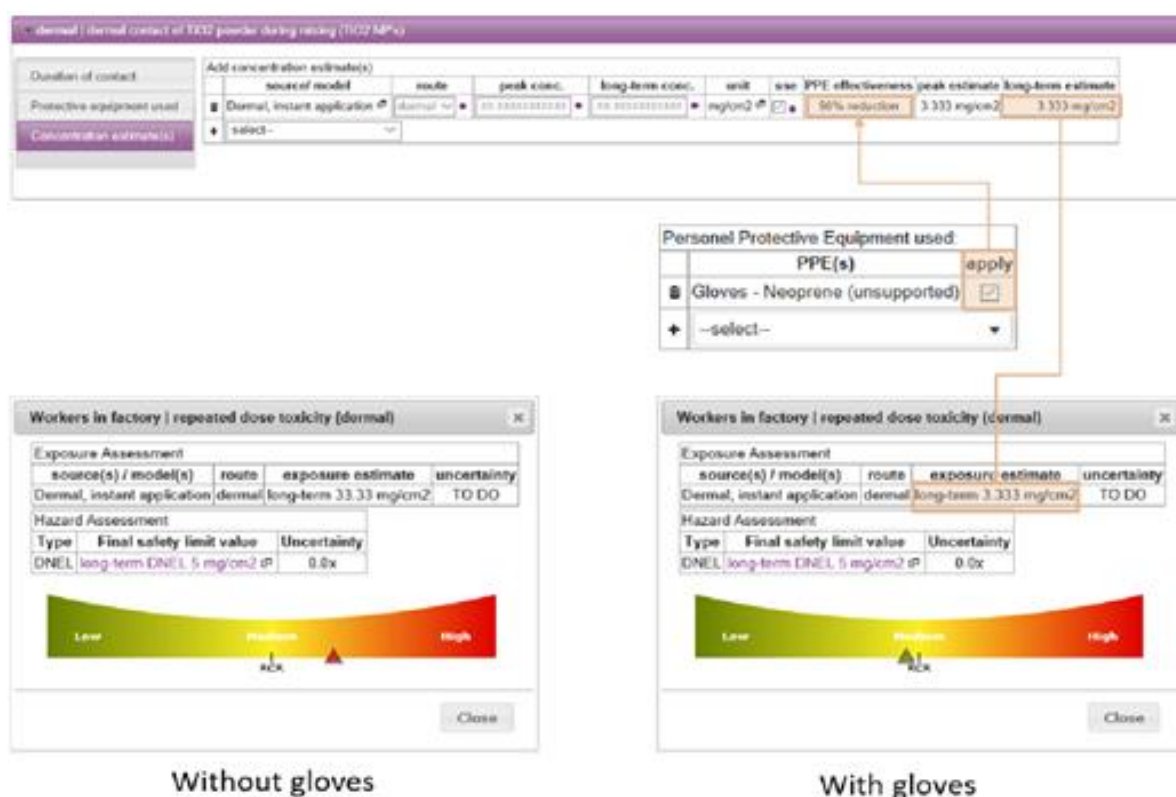
Figure 13: Effect of wearing gloves on occupational health risk associated with a given activity.

After clicking the "red-balance" in the overview, a dialog is shown presenting the estimated exposure value and the derived final safety limit value. A "risk ratio" is calculated dividing the "exposure value" with the "final safety limit value" indicated by the "red triangle" in case of a ratio > 1. The "blue triangle" is an indication of the risk ratio without adding potentially reducible uncertainty. This is an indicator for the user that it might be worthwhile to further reduce uncertainty. How to reduce uncertainty in the hazard assessment outcome depends on the uncertainty factors applied at the hazard and/or exposure assessment. Each applied uncertainty factor contains its "magnitude", "source" and an "advice text". The user can use this information to decide, if and what actions to take. An example of an action aiming to reduce uncertainty that one can think of is finding more relevant toxicological data (e.g., more similar

test material), or using higher tier exposure models or conducting actual concentration measurements. Actions to reduce the risk as such can also be considered (and are the only option if the blue triangle still indicates a high probability of risk). The user can choose to implement release mitigating measures or advise the use of personal protective equipment to mitigate the probability of risk (Figure 13).

GUIDEnano thus provides a complete risk assessment report including the data input and output of all modules. Integration into NanoCommons, and with the wide range of hazard and exposure models and datasets being made available via the KnowledgeBase, will further enhance the utility, accessibility and stakeholder acceptance and utilisation of the GUIDEnano tool, and is also an important step towards ensuring maximum return on EU-investment in nanosafety research.

GUIDEnano future development

GUIDEnano is still under development and continual work is performed to make the tools more user friendly according to the feedback of users. Nowadays, GUIDEnano work in the H2020 GRACIOUS project will lead to the incorporation of the GRACIOUS grouping and read across framework into the GUIDEnano framework. Another current development pathway, more specifically linked to the NanoCommons project, is to connect GUIDEnano to relevant exposure and hazard databases. A release activity database is already implemented in GUIDEnano and will be further populated to describe more activity especially during the use phase and the end of life of the NEP. Secondly, the connection with (eco)toxicological, NMs and PEP database (eNanomapper and NIST for NM data and PPEs building on existing databases, such as ECEL) is planned. However, some important technical aspects have to be figured out to perform this task, which will be discussed with Biomax and EwC to determine the optimal solutions.

Tools present in the NanoCommons framework and complementary to GUIDEnano will be added as a link in GUIDEnano. The user will have the possibility to use the full range of available tools, and in collaboration with the Training WP (WP9) of NanoCommons, some guidance on which tool to select for which specific NMs / organisms / exposure scenarios / compartments will be developed.

4. Dose-Response Modelling

Risk Assessment and Dose-Response Relationships

Dose-response modelling is a key step in the RA workflow, and its main aim is to characterize the relationship between the exposures and the observed adverse outcome [13] and eventually estimate human health guidance values, such as the DNEL value. In order to understand and model this relationship concerning toxic NMs as well as environmental factors, toxicology studies that involve multiple species, including humans and animals, are designed to investigate multiple exposure routes and media. Usually multiple dose-response curves are fitted, using *in vitro* and *in vivo* models, focusing on the dose response of the most sensitive organ, i.e. the target organ, of the species of interest for the exposure to the NM of interest. This gives information about the relationship between the actual dose that is delivered to the target tissue(s) and the biological effect(s) it causes deviating from the normal functions of the organism.

Traditionally, this analysis has been carried out differently for cancer and non-cancer endpoints on the grounds that cancer risks are assumed to be linear in the low-dose range that might be expected to occur in the human population [14] and can arise from exposure to any concentration of pollutant, whereas non-cancer effects are considered to exhibit threshold characteristics such that exposure below a certain concentration is associated with no appreciable risk of an adverse outcome. However, this distinction has become less pronounced, since our understanding of the respective health related processes has increased over time, giving rise to a more unified approach by combining the methods for cancer and noncancer health effects [15, 16, 17].

The NOAEL/NOEL approach

Quantitative methods, used to analyse cancer risks, involve fitting mathematical models to the observed tumour incidence data and extrapolating risk in a linear fashion to lower dose levels [14, 18]. On the other hand, quantitative risk assessment for noncancer effects was commonly based on determination of a No Observed Adverse Effect Level (NOAEL) from a controlled study in animals as an estimate of the threshold. In this context, the NOAEL is defined as the highest experimental dose that does not produce a statistically or biologically significant increase in adverse effects over those of controls. An "acceptably safe" daily dose for humans is then derived by dividing the NOAEL by a safety factor, usually 10 to 1,000, to account for sensitive subgroups of the population, data insufficiency, and extrapolation from animals to humans (see, for example, [19]). Alternatively, a Lowest Observed Adverse Effect Level (LOAEL) is used, which is defined as the lowest dose at which a significant adverse effect is detected. The NOAEL, or LOAEL in the absence of a NOAEL, serve as the point of departure (POD, also known as Reference Point (RP)) for the calculation of the Reference dose (RfD) through the application of uncertainty factors (UFs) designed to account for uncertainty in the data (e.g., lack of information about the differences in toxicokinetics and toxicodynamics between the test species and humans, differences in the calculation of the RfD among human individuals, or lack of knowledge in the database about other potential hazards for a particular chemical). The RfD term is used by the U.S. Environmental Protection Agency (EPA), while the Food and Drug Administration (FDA) uses the term allowable daily intake (ADI) and the Agency for Toxic Substances and Disease Registry (ATSDR) uses

minimum risk level (MRL). After application of the UFs, the RfD is presumed to represent a daily dose over a lifetime with no appreciable risk of adverse non-cancer health outcomes.

The BenchMark Dose (BMD) approach

Due to inherent limitations of the NOAEL/LOAEL approach, the European Food Safety Authority (EFSA) [21, 22] and the Joint FAO/WHO Expert Committee on Food Additives [23] have proposed the use of the Benchmark Dose (BMD) approach for deriving the RP used to calculate the margins of exposure (MOE) for substances that are both genotoxic and carcinogenic. For a more detailed explanation please see the EFSA report by Hardy et al. [24] (whose main concept is depicted in Figure 14). In the simplest case an experiment is conducted by exposing a predefined number of animals to different drug concentrations (including 0 for the so-called control unit) of the substance of interest in a predefined matrix (e.g. in air or water). After a predefined time, *in vivo* tests are conducted and selected indicators of effect are evaluated (e.g. number of deaths or weight of kidney). The resulting experimental dose-response results are fitted by two families of nested models of increased complexity, namely the exponential and the Hill families (Table 1). In order to select the best model within a family of models, more complex models must be compared to the corresponding simpler models in order to determine whether the addition of extra parameters significantly improves the model fit. This is done in a step-wise fashion until the most "optimal" (parsimonious) model has been selected. The selected model defines a concentration where a predefined value relative to the control group level of response is reached, which defines the start of an effect (e.g. 5% of death rate or 20% kidney weight increase).

The empirical probability distribution of BMD values associated to the selected BMR is calculated using bootstrapping; a large number of artificial data sets are generated, obtained by random sampling from a lognormal distribution with geometric mean (=median) defined by the fitted regression model, and geometric standard-deviations (GSDs) equal to the residual GDSs associated with the fitted model. Each artificial (bootstrap) data set is based on the original experimental design, i.e. the same dose groups and number of subjects within groups.

Table 1: continuous response models for the Exponential family and Hill family models.

Model	Number of model parameters	Model expression response (y) as function of dose (x)	Constraints
Continuous data			
Exponential family			
Model 1 [1)	1	$y = a$	$a > 0$
Model 2	2	$y = a \exp(bx)$	$a > 0$
Model 3	3	$y = a \exp(bx^d)$	$a > 0, d > 1$
Model 4	3	$y = a [c - (c-1)\exp(-bx)]$	$a > 0, b > 0, c > 0$
Model 5	4	$y = a [c - (c-1)\exp(-bx^d)]$	$a > 0, b > 0, c > 0, d > 1$
Hill family			
Model 2	2	$y = a [1 - x/(b+x)]$	$a > 0$
Model 3	3	$y = a [1 - x^d/(b^d + x^d)]$	$a > 0, d > 1$
Model 4	3	$y = a [1 + (c-1)x/(b+x)]$	$a > 0, b > 0, c > 0$
Model 5	4	$y = a [1 + (c-1)x^d/(b^d + x^d)]$	$a > 0, b > 0, c > 0, d > 1$
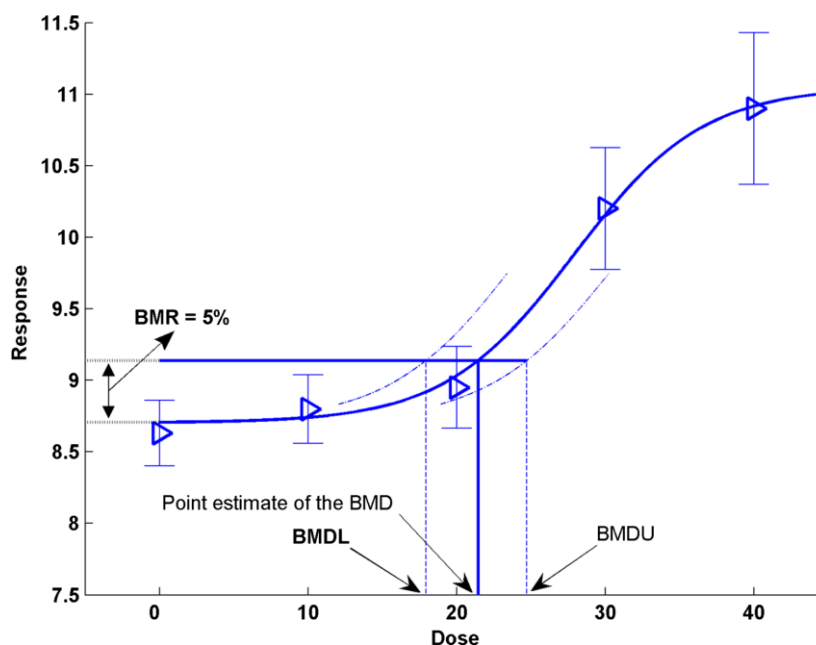
Figure 14: Key concepts for the BMD approach, illustrated using hypothetical continuous data. The triangles represent the observed mean responses, and are plotted together with their confidence intervals. The solid curve is a fitted dose–response model, which determines the point estimate of the BMD, defined as a dose that corresponds to a low but measurable change in response, denoted the benchmark response (BMR). The dashed curves represent, respectively, the upper and lower 95% confidence bounds (one sided) for the effect size as a function of dose. Their intersections with the horizontal line are at the lower and upper bounds of the BMD, denoted BMDL and BMDU, respectively. It should be noted that the BMR is not defined as a change with regard to the observed mean background response, but with regard to the background response predicted by the fitted model. In the Figure, the BMD corresponds to a 5% change in response relative to background (BMR = 5%) (Figure and legend slightly modified from [24]).

Comparison between the NOAEL and the BMD approaches

A direct comparison between the NOAEL and the BMD approach is provided in references [15, 20] and is summarised in Table 2 [20]. The NOAEL/LOAEL approach for the determination of PODs has been associated with well documented limitations. Most importantly, NOAEL is based upon the study design, sample sizes, dose spacing, background levels, the power of the design (and statistical test) and on a hypothesis testing approach in that failure to reject the null hypothesis is taken as evidence of no difference. The latter differs from the conventional view of the interpretation of the failure to reject the null hypothesis. As a result, the smaller the experiment or the more variable the endpoint the less chance of detecting a real effect exists. Also, the NOAELs tend to be higher for measures with a high control/background level because it is more difficult to demonstrate a statistically significant difference in standard designs than for measures with low control/background levels. Essentially, NOAEL is a single dose level and takes no account of the dose-response relationship in the data and since it is not an estimate of a dose, it cannot be presented with a measure of precision such as a confidence interval (CI). As it is a single experimental dose level, it is not necessarily representative of the true "threshold". Thus, NOAEL cannot be considered a risk or response-free exposure level and is not necessarily a dose where there is no effect (i.e. below a threshold) [17].

On the other hand, the main limitation of the BMD approach is that it is a more complicated process, is time consuming, and requires a large number of data points covering a wide range of dose quantities. The number of samples available affects the calculated BMDL and NOAEL values. For a small sample, the calculated BMDL can be bigger or smaller than NOAEL. When test sample size (i.e. no. of tested animals) is very big, BMDL is higher than NOAEL.

Despite the limitations of the BMD approach, it has become the US EPA's preferred dose–response assessment method. Other authorities such as EFSA also use the BMD method for food safety RA [25,26].

Table 2: Advantages and limitations of the NOAEL and BMD approaches [20]**.**

		BMD	NOAEL
Advantages		Not limited to experimental doses	Can be used when data is not amenable for BMD modelling
		Less dependent on dose spacing	Easy to derive
		Appropriately accounts for variability and uncertainty resulting from study quality	Has been the standard method for deriving a POD for decades (e.g., is familiar to most risk assessors)
		Takes into account the shape of the dose–response curve and other related information	
		Corresponds to consistent response level and can be used to compare results across chemicals and studies	
		Flexibility in determining biologically significant rates	
Limitations		Ability to estimate BMD may be limited by the format of data presented	Highly dependent on dose selection
		Time consuming	Highly dependent on sample size
		More complicated decision-making process	Does not account for variability and uncertainty in the experimental results (e.g., does not account for study quality appropriately)
			Dose–response information (e.g., shape of dose–response curve) not taken into account
			Does not correspond to consistent response levels for comparisons across studies
			A LOAEL cannot be used to derive a NOAEL

Software implementations of the BMD approach

The BMD approach has been implemented in two software packages, namely the Benchmark Dose Software (BMDS) developed by EPA in the US and the PROAST software developed by RIVM. RIVM and EPA aim to achieve consistency between the BMDS and PROAST software, but there are still some differences, including a number of default settings for statistical assumptions. Furthermore, the two software packages differ in functionalities [24]. Examples of useful functionalities in PROAST are the possibility of statistically comparing dose-responses among subgroups (covariate analysis), and the larger flexibility in plotting. Currently, two web applications of PROAST are available, which do not

include all functionalities of the R package, but only the usual dose-response analysis of toxicity data can be done. These web tools are provided by EFSA and the RIVM team (https://efsa.openanalytics.eu/, https://proastweb.rivm.nl/).

IntPROAST R package and workflow

Given the importance of the BMD approach in the RA workflow, we have integrated the PROAST software for continuous dose-response into the NanoCommons infrastructure through the Jaqpot computational platform and in particular through the IntPROAST R package web implementation of the software. As a first step, the user is requested to upload the dose-response data, indicate the name of the dose variable, the BMR percentage value and the number of bootstrap samples generated in order to calculate BMD confidence intervals. For the latter two parameters there are default values given. The output consists of a table with log-likelihood values for all models considered from the two families of models, as well as associated graphs and the print out on screen, as generated by the PROAST package. Particularly, input data should include the following:

1. Dose-response matrix in JSON format with compulsory column names: id, dose, response
2. Indication of the name of the dose variable in the above data table
3. BMR percentage (default 5%)
4. Number of Bootstraps (default 10,000)

Output data are returned as a single array in JSON format including:

1. A table with all log-likelihood values also including the number of parameters, the name of the model and the model's family name
2. A text explaining the results as generated by the PROAST software
3. All fitted graphs of the models saved as PNG files in the working directory and also encoded in base64 scheme within the JSON file.

IntPROAST includes the following steps:

1. Generate a PROAST-like R data environment able to store the input data
2. Call the PROAST algorithm with the generated environment in order to:
    i. Calculate the geometric mean at each tested dose
    ii. For each family member (model) starting from the one with fewest input parameters
        1. Fit the model with the means
        2. Calculate the log-likelihood
    iii. For each model within the family of models compare the log-likelihood with the preceding model's log-likelihood, if the difference is higher than the corresponding critical difference keep the current model as the family representative
    iv. For each selected family representative model
        1. Test the goodness of fit by comparing its log-likelihood to that of the full model; if the distance is higher the model is rejected. If all family representative models are rejected, additional data are necessary.

Below we present the analysis results when employing IntPROAST to determine the BMD using an artificial example data set uploaded to the eNanoMapper database [27]: https://apps.ideaconsult.net/ambit2/dataset/R545.

The dose variable id is the following feature:

https://apps.ideaconsult.net/ambit2/feature/22127

and the response variable is:

https://apps.ideaconsult.net/ambit2/feature/22137

A screenshot of the data can be seen in Figure 15. Figure 16 shows the output of the IntPROAST R package within the R environment, where both input and output files are in the JSON format.

```
"https://apps.ideaconsult.net/ambit2/feature/22127"
"https://apps.ideaconsult.net/ambit2/feature/22137"
"https://apps.ideaconsult.net/ambit2/feature/22200"
"https://apps.ideaconsult.net/ambit2/feature/22213"
"https://apps.ideaconsult.net/ambit2/feature/22252"
"1"  0.268 0.667 -5.331 2 0.352
"2"  0.115 0.759 -5.342 1 0.213
"3"  0.173 0.596 -5.331 1 0.322
"4"  0.088 0.828 -5.334 0 0.354
"5"  0.309 0.635 -5.335 1 0.264
"6"  0.168 0.836 -5.346 1 0.168
"7"  0.158 0.802 -5.328 0 0.354
"8"  0.207 0.743 -5.328 4 0.354
"9"  0.122 0.969 -5.367 0 0
"10" 0.28 0.759 -5.343 1 0.264
"11" 0.199 0.81 -5.345 2 0.189
"12" 0.311 0.645 -5.317 3 0.452
"13" 0.249 0.705 -5.341 2 0.234
"14" 0.206 0.623 -5.326 1 0.354
"15" 0.134 0.726 -5.336 2 0.271
"16" 0.213 0.823 -5.346 0 0.25
"17" 0.252 0.749 -5.328 3 0.354
"18" 0.271 0.783 -5.332 2 0.354
"19" 0.336 0.541 -5.297 4 0.65
"20" 0.255 0.817 -5.339 2 0.234
"21" 0.284 0.638 -5.329 0 0.334
"22" 0.281 0.634 -5.308 3 0.55
"23" 0.351 0.746 -5.318 2 0.516
```

Figure 15: Screenshot of the dose-response data set.
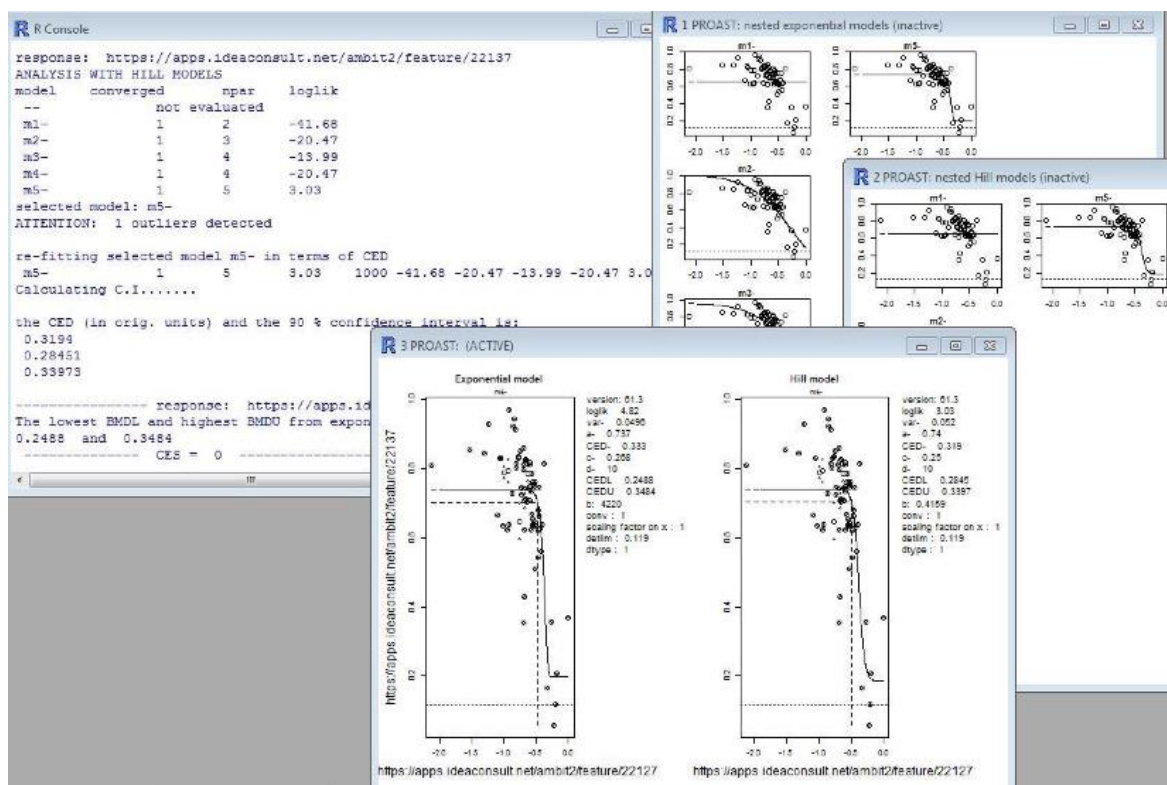
Figure 16: Screenshot of the dose-response dataset analysis using the IntPROAST package. In the R Console the log-likelihood values of the various models are shown together with a short explanation of the output. Graphs show all model fits considered and also the best fit models per family.

Future work: PROAST/Plumber API

NTUA is currently developing an additional PROAST integration application as a REST API based on 'plumber' R Cran library exposing just one POST endpoint [28]. Plumber library offers a simple solution to expose R code as a service, just by adding in the R code special annotations defining the service endpoint(s). This implementation uses the standard JSON format for posted input and rendered output since it depends on the 'jsonlite' R Cran package, thus enabling seamless integration with Jaqpot as well as any other platform. The latter is further facilitated by the containerization of the tool that empowers the stability of the application, the transferability regardless of the running environment, the elimination of application conflicts and the scalability to accommodate any kind of user needs. The implementation follows the same workflow as the IntPROAST R package, which has been described above, and the current version as well as future updates will be available at GitHub [https://github.com/KinkyDesign/proast.git].

5. Read across methodologies

Predictive modelling is an *in-silico* approach that can be used for hazard assessment of NMs, where there is an absence of experimental toxicology data. Predictive models can produce reliable estimations of toxicity related end-points much faster and with reduced cost compared to *in vivo* methods and most importantly do not involve animal testing. Over the past few years, the nanosafety community has encouraged the development of such alternative non-testing methods for the toxicological investigation of NMs. These computer-aided methods aim to contribute to the prioritization of NMs for safety evaluation and to support the regulatory decision-making. One successful approach [29] is the adaptation of the quantitative structure-activity relationship (QSAR) modelling methodologies to the special requirements of NMs, which arise due to their complex structures. The produced models are presented in the literature as nano-QSARs or QNARs (quantitative nanostructure-activity relationship) models. However, in order to ensure the functionality of the QNAR approaches, sufficiently large (more than 20 distinct NM samples) and diverse datasets should be provided [30].

The European Chemicals Agency (ECHA) through the Read-Across Assessment Framework (RAAF) has introduced the alternative read-across non-testing strategy [31, 32], for the prediction of NM toxicity, especially in the absence of sufficiently large datasets for the development of reliable QSAR models [30]. The read-across concept is based on the empirical knowledge that similar materials may exhibit comparable properties thus, the estimation of the hazardous effects of non-tested NMs can be achieved using data within a group of comparable NMs [30, 33, 34].

The entire proposed strategy consists of seven well defined steps,[35] as follows:

1. Determination of the structural characteristics of NMs (composition, including surface chemistry and any impurities, size, shape etc.).
2. Development of an initial grouping hypothesis that correlates an endpoint (e.g. a toxicity index), to different behaviour and reactivity properties, including solubility, zeta potential, dispersibility, hydrophobicity, dustiness, biological activity (redox formation, gene expression etc.), photoreactivity etc. [35, 36, 37]. Assignment of the samples to groups.
3. Gathering of the above properties (depending on the grouping hypothesis) for each NM.
4. Construction of a data matrix including properties and endpoints.
5. Assessment of the applicability of the approach using computational techniques (e.g. Principal Components Analysis (PCA), hierarchical clustering [35, 36], random forests [35], linear discriminant analysis (LDA) [38] etc.), and data gaps filling. If no regular pattern emerges, an alternative grouping hypothesis must be proposed (step 2).
6. In case that the grouping hypothesis is robust, but adequate data are not available, additional testing should be considered to complete the datasets.
7. Justification of the method.

There are two approaches regarding the read-across framework, supported by ECHA and OECD; the analogue and the category/grouping approach.

In the analogue approach the prediction is limited to a small area of the data space; one source NM can be used for the endpoint estimation for a single or more target NMs, or two or more source NMs can be used to make predictions for a single or several target NMs. The read-across methodologies apply an interpolation strategy "locally" among similar samples which, depending on the provided data

-numerical or discrete-, can be quantitative or qualitative [36]. The methods for the prediction of each endpoint range from simple average value calculations, or simple linear interpolations to more complicated methods applying QSAR methods locally (e.g. *k*-nearest neighbours, partial least squares, random forests) [30, 39].

In the categorical approach, the NM samples are organized into groups of similar compounds. Groups are formed considering structural similarities between samples, and it is assumed that due to these similarities, the biological or toxic activity of the NMs within a group follows a regular pattern. Groups of NMs can be further divided into subgroups based on interdependencies in nanodescriptors and the formation of these subgroups can be "tuned" in order to gain satisfactory predictions [38, 40]. Other studies have investigated alternative grouping possibilities including PCA, LDA,[38] two-dimensional hierarchical clustering [36] or considering the NM mode-of-action [37]. For the estimation of the endpoint of a target NM in a group, the analogue approach can be applied.

Several read-across tools and methods for the preliminary hazard assessment of NMs have been proposed in the literature [33]. Gajewicz *et al.* [30] proposed a novel quantitative read-across (QRA) approach for data gap filling of NMs using the one-point-slope, the two-point formula and the equation of a plane passing through three points. Their Nano-QRA model proved to have high predictive capabilities, when tested with the same dataset used by Puzyn *et al.* [41]. Helma *et al.* [39] introduced recently the nano-lazar framework for NM read-across predictions. The similarity levels for the selection of neighbours is based on Tanimoto/Jaccard index and on weighted cosine similarity. Three local regression algorithms are available; weighted local average, weighted partial least squares regression and weighted random forests. Helma et al. [39] tested the performance of their methods using the dataset initially presented by Walkey *et al.* [42] consisting of 121 gold and silver NMs that are characterized by physicochemical descriptors, the protein corona fingerprints (PCF) and by MP2D fingerprints calculated for core and coating compounds with defined chemical structures. They reported R^2 values equal to 0.68, for the prediction of the cell association with human A549 cells, using only the protein corona fingerprints and the weighted random forest algorithm, in a 10-fold cross validation scheme. Varsou *et al.* [40] presented the toxFlow web application, which integrates physicochemical, omics and biology information data for read-across toxicity prediction of ENMs. Neighbor selection is based either on the cosine similarity between NMs or a distance metric (Euclidean, Manhattan). Using only the gold NMs of the Walkey *et al.* [42] study and performing enrichment analysis to the PCF data prior to read-across, Varsou et al. reported R^2 values 0.97 in the toxicity prediction, by employing a weighted average algorithm and a leave-one-out validation scheme. Very recently, Lamon *et al.* [33] and Aschberger *et al.* [43] presented case studies applying the ECHA RAAF framework for the read-across prediction of hazard endpoints of nanoforms of TiO_2 and of Multi-Walled Carbon NanoTubes (MWCNTs) respectively. The first of these studies has been released as an official OECD document [44].

As described previously, the development of a read-across scheme typically assumes a hypothesis, which is evaluated and assessed in terms of its adequacy to fill data gaps. The read-across hypothesis involves both the selection of the most informative descriptors that can predict the endpoint of interest and the definition of the source NMs, that can be considered as neighbours to the target NM. This procedure is iterated in a trial-and-error fashion until a hypothesis producing successful read-across predictions is determined. The procedure is time-consuming and due to the complexity of the problem, it does not guarantee that the produced read-across model is optimal.

NanoCommons read-across methodologies

Read-across method based on mathematical optimization

In the context of the NanoCommons project, NTUA has developed a robust and reliable read-across method that automates the ECHA RAAF for the prediction of NM undesired properties, by focusing on two separate goals: First, the reduction of the available dataset, by removing the instances (variables) that add noise rather than useful information to the analysis. Second, the definition of the neighbour boundaries which indicate the source NMs that are considered similar to the target NM. These two different goals are achieved simultaneously through the development of a Mixed Integer NonLinear Programming (MINLP) problem, where the objective is to minimize the mean squared error (MSE) between the experimental values and the produced predictions with respect to selecting the most informative descriptors and defining the neighbour boundaries.

Available data

The methodology assumes the availability of a dataset containing the values of L descriptors and the end-point for N NMs. The data are first scaled using a standardization (e.g. Gaussian normalization) or a normalization (e.g. min-max) method, so that descriptors with substantially different numerical ranges contribute equally after the transformation to the overall analysis [45]. The dataset is denoted by $\{\mathbf{x}_i, y_i\}$, i=1, ..., N, where $\{\mathbf{x}_i\}$ is a vector containing the values of the L descriptors for the ith NM and y_i is the endpoint value of the ith NM.

Set of variables

The main outcomes of the MINLP problem are:

$attr_l$: a binary variable indicating if the descriptor l is selected, l=1, …, L.

thr: a continuous variable that defines a threshold for the selection on neighbouring NMs. If the Euclidean distance between two NMs is equal or less than thr, these two NMs are considered as neighbours.

A number of additional variables are used for the construction of the MINLP problem:

$dist_{i,j}$: a continuous variable containing the Euclidean distance between NMs i and j, i=1,…,N, j=1,…,N

$neib_{i,j}$: a binary variable taking the value of 1 if NMs i and j are neighbours and 0 if they are not, i=1,…,N, j=1,…,N

$pred_i$: a binary variable taking the value of 1, if NM i has at least one neighbour and 0 if it has no neighbours, i=1, …, N

$\tilde{y}_i$: a continuous variable containing the predicted endpoint read-across value for the ith NM, i=1, …, N

Mathematical formulation

Set of constraints: Equation 1 computes the Euclidean distance between all pairs of NMs taking into account only the selected descriptors:

$$dist_{i,j} = \sqrt{\sum_{\ell=1}^{L} attr_\ell \left(x_{i,\ell} - x_{j,\ell} \right)^2} \quad , i = 1,\ldots,N, \quad j = 1,\ldots,N, \quad i \neq j$$

$$(1)$$

Equations 2-4 ensure that the two NMs are considered as neighbours only if their Euclidean distance is lower than the threshold. In this case the binary variable takes the value of 1, otherwise the value of 0 is assigned to this variable. In Eqs. (3), (4) m is a small number (equal to 10^{-3}):

$$neib_{i,j} \geq m \cdot (thr - dist_{i,j}), \; \forall \; i, j \in \{1,\ldots,N\}, \; i \neq j \qquad (2)$$

$$1 - neib_{i,j} \geq -m \cdot (thr - dist_{i,j}), \; \forall \; i, j \in \{1,\ldots,N\}, \; i \neq j \qquad (3)$$

$$neib_{i,i} = 0, \; \forall \; i \in \{1,\ldots,N\} \qquad (4)$$

Equation 5 computes the read-across predictions as weighted averages of the endpoint values of neighbour NMs:

$$\hat{y}_i = \frac{\sum_{j=1}^{N} y_j \cdot \frac{neib_{i,j}}{1+dist_{i,j}}}{\sum_{j=1}^{N} \frac{neib_{i,j}}{1+dist_{i,j}}}, \forall \; i \in \{1,\ldots,N\} \qquad (5)$$

For NMs without any neighbour, the prediction is obviously equal to 0 and it is not accepted. An additional constraint is used to guarantee that the percent of NMs with at least one neighbour for which an endpoint prediction is produced, is greater than or equal to a predefined percentage denoted by *predFactor* as shown in Equations 6-8:

$$\sum_{i=1}^{N} pred_i \geq predFactor \cdot N \qquad (6)$$

$$pred_i \geq neib_{i,j}, \forall \; i \in \{1,\ldots,N\}, \forall \; j \in \{1,\ldots,N\} \qquad (7)$$

$$pred_i \leq \sum_{j=1}^{N} neib_{i,j}, \forall \; i \in \{1,\ldots,N\} \qquad (8)$$

Objective function: The objective function to be minimized is the MSE between the endpoint read-across predictions and the actual endpoint values over all the NMs with at least one neighbour (Equation 9).

$$min \frac{1}{\sum_{i=1}^{N} pred_i} \sum_{i=1}^{n} pred_i (y_i - \hat{y}_i)^2 \qquad (9)$$

Due to the complex structure of NMs, different types of data and descriptors are often used for NM characterization. These may include physicochemical, biological, quantum-mechanical, image, biokinetics descriptors etc. In NTUA's toxFlow application [40], the use of two similarity criteria for defining different thresholds was demonstrated for selecting the neighbours if different types of data are available. In this approach, distances can be calculated between all substances separately for the different types of data, and two NMs are considered as neighbours if both distances are lower than the corresponding thresholds. Based on this idea, the MINLP formulation described before can be extended

to account for multiple similarity criteria, thus increasing the confidence in the read across predictions.

Read-across method based on the *k*NN methodology

NovaM has developed a read-across method based on the *k*-nearest neighbour (*k*NN) machine learning algorithm. *k*NN can be considered as a read-across strategy [46], as it requires experimental observations of only a few neighbours (similar NMs) to the query NM, in order to compute the endpoint prediction. The *k*NN methodology is a "lazy" learning technique, that classifies an instance based on the majority vote of the *k* closest training examples (neighbours). In case that the endpoint has a numeric class, the prediction is the distance weighted average of the endpoint of the selected neighbours. An optimal *k* value is selected based on the calculated Euclidean distance between all instances and as weighting factors the inverse distances are used [47, 48]. In the case of a categorical endpoint each instance is assigned to the class indicated by the weighted majority vote of the *k* closest neighbours.

Another important aspect of the analysis -apart from the simple endpoint prediction- is the possibility to observe the groups of *k* neighbours of each test NM, and therefore to specify and map the analogous space (Figure 17), which is a prerequisite of the read-across framework,[31] and which can be used to support the justification of the read-across hypothesis.
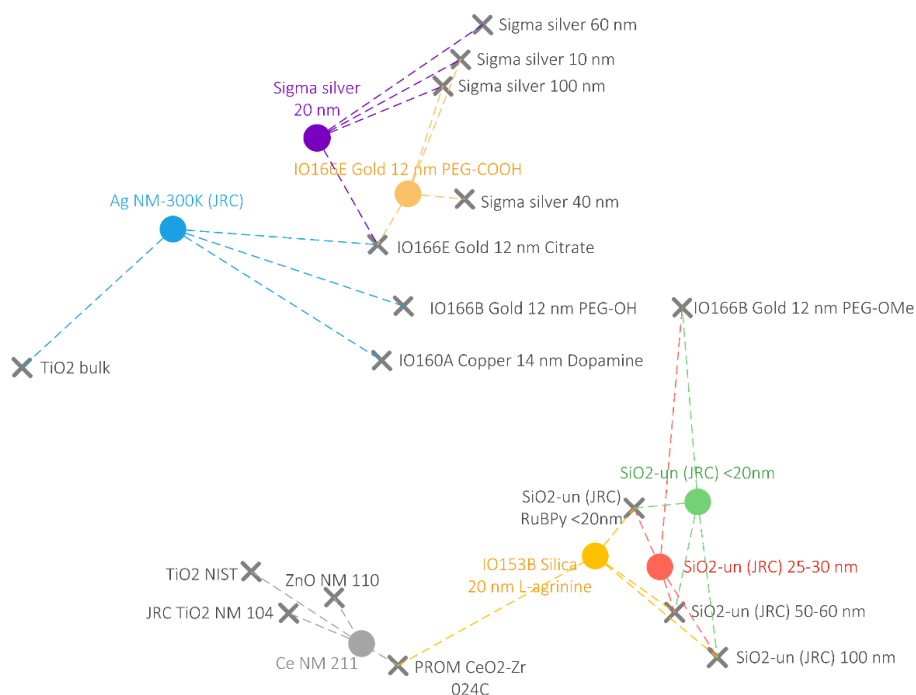
Figure 17: Example of a qualitative representation of the neighbouring space of the training and the test NM set utilising a set of descriptors determined experimentally and computationally. Test NMs are depicted with coloured circles, whereas training NMs are illustrated with grey crosses. The four selected-closest neighbours for each test NM are defined via dashed lines.

Future work

After the mathematical formulation of the two read-across methodologies, efforts are now focused on their validation using test data sets, their implementation as web services and their integration into the NanoCommons knowledge infrastructure. The methods will be offered openly to the scientific community via user-friendly interfaces, which will support their integration into risk assessment workflows and will facilitate the computer-aided design of novel NMs. To achieve this goal, the methods will be developed using open source platforms (e.g. KNIME platform) or programming languages (e.g. R). The web services will be hosted on the platforms of Jaqpot (NTUA) and Enalos Cloud (NovaM) and will be used within a risk assessment framework for NMs under the auspices of NanoCommons project. One additional new read-across approach is being developed by NTUA. This method is following the categorical approach and uses again mathematical programming to precisely partition the available training NMs into groups of similar behaviour. All read-across approaches developed within the NanoCommons project will be presented in full detail in Deliverable Report D6.2.

6. Biokinetics models

PBPK modeling

Biokinetics offer a methodology for predicting the internal distribution and exposure of a NM in an organism, which can be of particular importance in an RA workflow. Compartmental modelling is a concept broadly used in pharmacokinetics for describing the biodistribution of a substance inside an organism. Physiologically-based pharmacokinetic (PBPK) models represent one of the two major approaches used in compartmental modelling, with empirical models being the second one. PBPK models are mechanistic; they consist of compartments representing real organs and tissues, whose number varies based on the target substance, species, administration route and available information. A common approach is to incorporate in the model the main body tissues, i.e. brain, heart, kidney, skin, spleen, liver, lung, gut, bone, adipose and muscle [49]. Nevertheless, the dimensionality of a PBPK model can be reduced using lumping methods [50, 51]. In most cases, PBPK models are utilised for describing the kinetics of a substance in the whole body of a species, thus such models are more formally called "whole body physiologically-based pharmacokinetic models" (WBPBPK). However, there are some models developed to describe in detail the kinetics of an organ, which is divided into separate sub-compartments. This modelling approach is called "partial" PBPK models [52].

WBPBPK can be grouped into two classes: perfusion-limited and permeability-limited models. The first group applies for small lipophilic molecules, where the limiting step of substance absorption is the transportation to the tissue via the blood flow. On the other hand, permeability-limited models assume that the limiting process is the membrane permeability which acts as a diffusion barrier, so it applies for larger and hydrophilic molecules [53], which is the case for NMs. The compartments of a WBPBPK model are interconnected through the arterial and venous blood pools; all organs receive blood from the arterial blood compartment and the outflow ends up in the venous blood compartment. The blood loop is closed by the lungs compartment, in which the blood flow is reversed in comparison to the rest of the organs [54]. Apart from the regional blood flows, the underlying physiology also defines the second set of substance-independent parameters, the organ volumes. Besides these physiological parameters, PBPK models incorporate information about the target substance as well, through the substance-dependent parameters, which can be obtained using data generated by *in vivo*, *in vitro* or *in silico* experiments [55].

PBPK models have inherent advantages due to their mechanistic nature. Firstly, they enable predictions of concentration/mass profiles of individual organs and not just of plasma. In addition, their relation with physiology and modularity facilitate the integration of literature information, making predictions prior to *in vivo* experiments possible [56]. Lastly, their biggest advantage is the ability to perform inter-species (e.g. from rat to human) or intra-species (e.g. from adults to children) extrapolation through scaling methods.

Nano-PBPK models
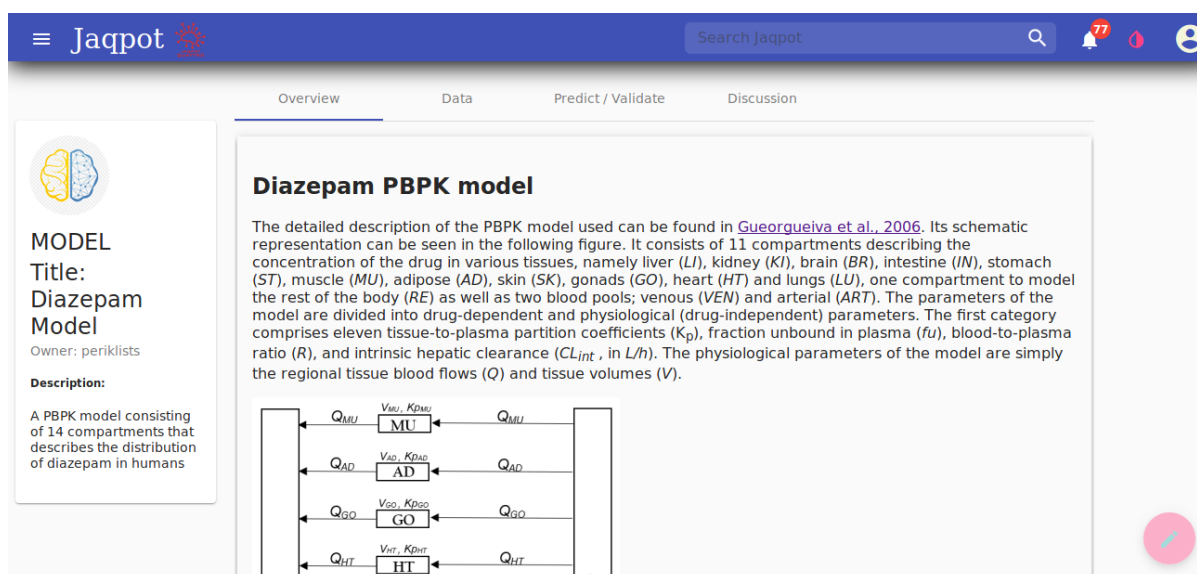
Several PBPK models that describe the biodistribution of NMs can be found in literature and an extensive review of the modifications of traditional PBPK models that need to be considered in the case of NMs can be found in [58,59]. A list containing the most notable nano-PBPK models is presented in Annex 3. Note all those models are accompanied by the data used for the model development, making

the models reproducible.

From these efforts, the model of Li et al. [59] stands out for introducing the concept of phagocytizing cells (PCs) via incorporation of a relevant sub-compartment in each organ. This compartmental modelling of the process of NM uptake by PCs registered satisfactory results based on the goodness-of-fit of the models on the experimental data. Carlander et al. [60] further built on this model by proposing some minor modifications that generalised the model, making it suitable for analysing the biodistribution of non-degradable NMs intravenously injected in rats. Specifically, they demonstrated the ability of this modified model to adequately describe the distribution of pegylated polyacrylamide, uncoated polyacrylamide, gold and titanium dioxide NMs in the body of rats.

Jaqpot biokinetics infrastructure

NTUA has developed all the necessary infrastructure to develop, host and share PBPK models through the Jaqpot computational platform. So far, the platform has been tested for simple molecules. The Jaqpot functionalities will be briefly demonstrated via a PBPK model for the diazepam drug. The first component of the web tool is an overview of the model in Markdown language, which includes a detailed verbal and schematic description of the model, as well as a link to the original study. Figure 18 presents a screenshot of the 'Overview' tab.

Figure 18: Screenshot of the 'Overview' tab of the PBPK web tool.

The second utility offered by the web tool concerns the dependent and independent features. Particularly, the model creator, via the data tab, can add a short description, units and ontological class for each of the features of the model, as illustrated in Figure 19.
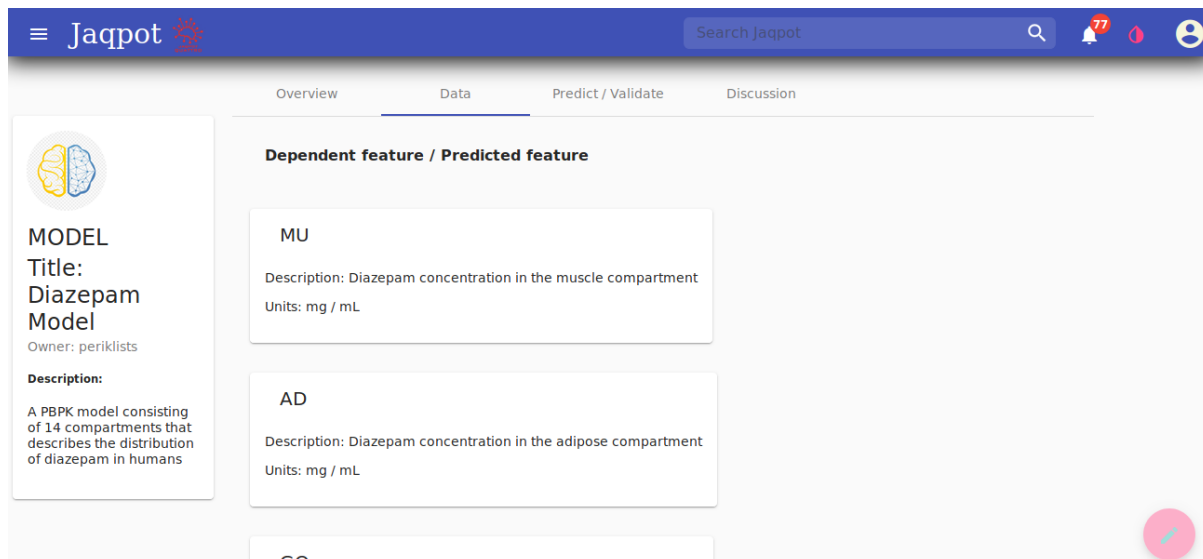
Figure 19: Screenshot of the 'Data' tab of the PBPK web tool.

The main functionalities of the Jaqpot PBPK web tool are provided through the 'Predict' tab (Figure 20). The user can upload a dataset containing the values of the independent features or manually enter the values. More specifically, for the showcase model the user can insert the physiological parameters of the specific individual where the drug is administered (gender, weight), drug related information (dose, infusion time, initial concentration in each of the 11 compartments, i.e. at time zero from which the model calculates the concentration at any time-point greater than 0), and the duration and time step that will be used in the simulation.
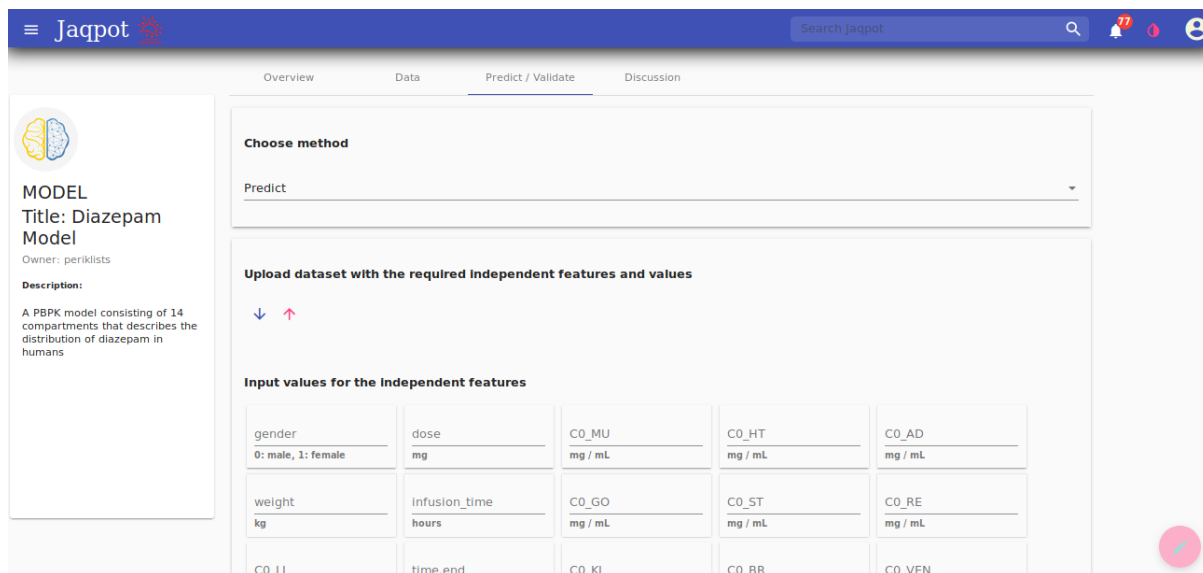
Figure 20: Screenshot of the 'Predict' tab of the PBPK web tool.

When finished, the user can click on the start button, initiating the prediction process. Shortly after, the mass profiles of the diazepam in the different organs are generated and presented in a table format, as seen in Figure 21.
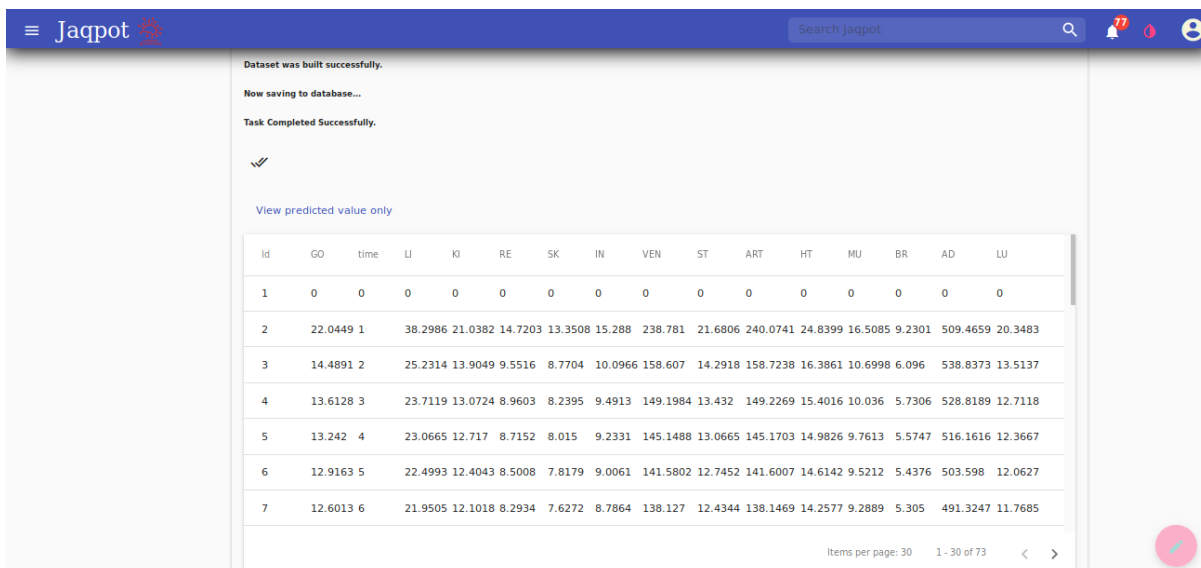
Figure 21: Screenshot of the table-formatted mass profiles of the Diazepam model.

The results can be downloaded in csv format and be further processed to make inference offline. Figure 22 shows an example of the offline post-processing, depicting the drug mass profile in the lungs.
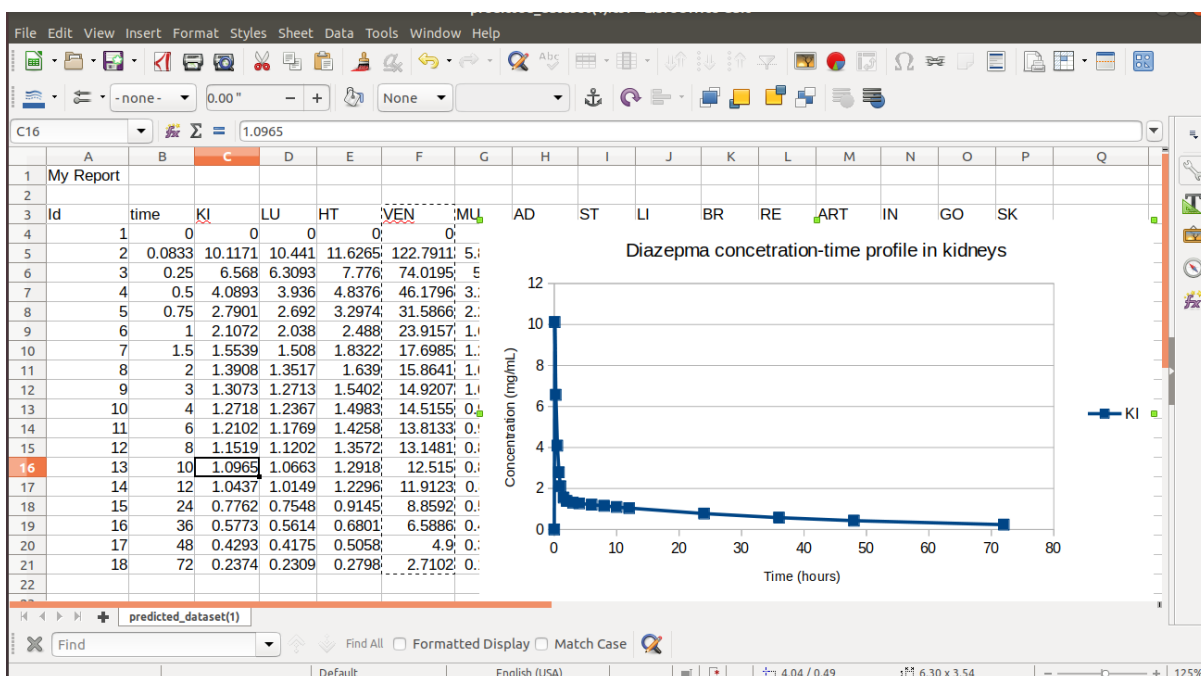
Figure 22: Example of offline usage of the csv-formatted results.

Lastly, a 'Discussion' has been incorporated in the web tool, which offers the opportunity to add comments and ask questions about the uploaded PBPK model (Figure 23).
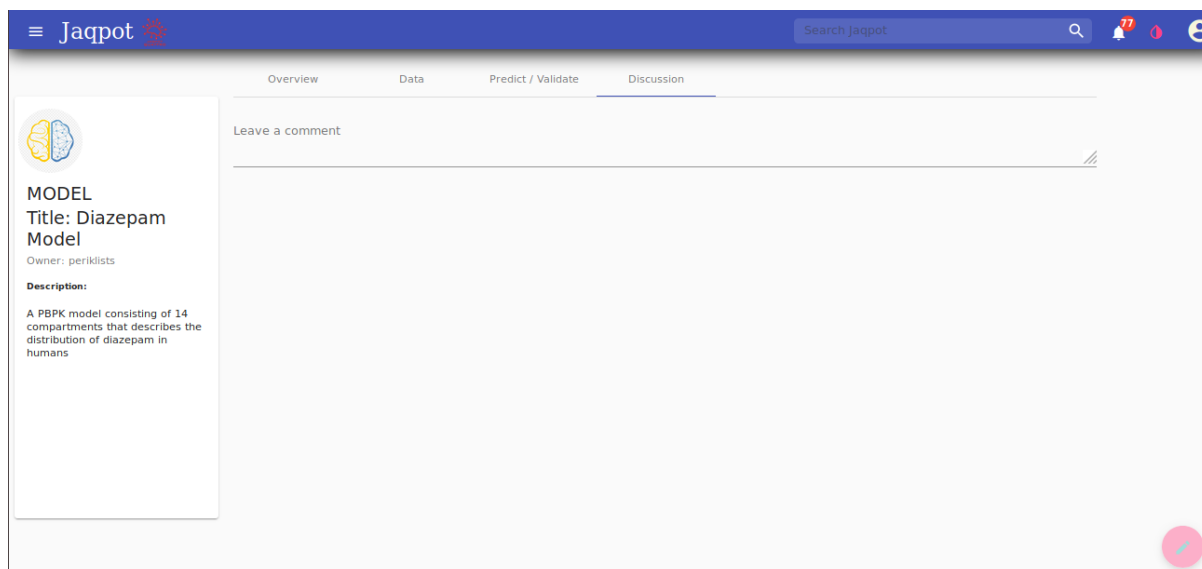
Figure 23: Screenshot of the 'Discussion' tab of the PBPK web tool.

Future work

One important goal for further development of the Jaqpot PBPK modelling infrastructure is to provide means for creating and hosting practically any PBPK model. NTUA will create an R library for the deployment of custom PBPK models by R users. This R package will support the upload of a PBPK model on the Jaqpot server by providing the names of the dependent and independent features, as well as the differential equations and other functions, if applicable (e.g. covariate models for organ scaling). The Jaqpot Graphical User Interface (GUI) will be extended by providing functionalities to automatically produce concentration-time profiles in the Jaqpot environment. This improvement will enable generation of mass/concentration time profile plots on the UI, thus minimising the need for offline post-processing.

NTUA will then integrate a number of nanoPBPK models in the Jaqpot system with the goal of creating a library of such models, similar to the nanoQSAR model repository reported in Deliverable report D5.4. This development will be supported by a Transnational Access (TA) project, as Prof. Dingsheng Li (University of Nevada), the first author of publication [59] has applied for a TA with the goal of creating a Jaqpot implementation of the PBPK model presented in [59], which describes the biokinetics of polyethylene glycol-coated polyacrylamide NMs in rats.

While the PBPK model doesn't currently consider the adsorption of a protein corona around NMs, and interesting area for further work would be to determine whether it is possible to include prediction of corona in blood (e.g. using partner UCD's protein corona model described in Deliverable report D5.6 - First corona simulation tools integrated into NanoCommons KnowledgeBase) into the PBPK model. NTUA and UCD will explore this possibility.

7. Bayesian networks

Bayesian networks and risk assessment for NMs

Bayesian networks are excellent tools for applications were data integration and reasoning under uncertainty are needed. This is the reason why they have been utilised in estimating risk of NMs, along with methodologies like stochastic multicriteria acceptability analysis (SMAA), Weight-of-Evidence (WoE), grouping, QSAR models, as well as combinations of the aforementioned methods. Wiesner et al.,[61} and Marvin et al.,[62] were among the first to recognise the suitability of Bayesian networks in risk forecasting of NMs, highlighting their advantages. An early application of Bayesian networks was that of Money et al. [63], who developed a Bayesian network for modelling the risk of silver NMs exposure in aquatic environments. In a later publication they updated the initial model, which was then validated and used for sensitivity analysis [64]. Marvin et al. [62] developed a Bayesian network for hazard ranking of metal NMls (specifically they used data for TiO_2, SiO_2, Ag, CeO_2, ZnO), using as nodes the physicochemical characteristics of a NM, the biological effects and exposure routes. On a similar wavelength, Murphy et al. [65] deployed a Bayesian network using Ag, TiO_2 and CNTs data and combined physicochemical properties and exposure potential to conclude on a RAt via a hazard node.

A different approach of using Bayesian networks with the goal of risk assessment of NMs is the development and quantitative analysis of AOPs and AOP networks [66]. An AOP is a representation that begins with a biological perturbation caused by a molecular initiating event (MIE), which is causally linked via a series of key events (KE) to an adverse outcome (AO), happening at a biological level of organization that is related to RA [67]. To the best of our knowledge, there has been only one implementation of Bayesian networks on AOPs: Jeong et al. [68] developed an AOP for relating the reproduction failure of *Caenorhabditis elegans* through oxidative stress caused by Ag NMs and causal relationships between the building blocks of the AOP were established using Bayesian networks.

Structure and properties

Bayesian networks constitute a robust tool for investigating causal relationships between variables and for making predictions. One of their key features is the ability to integrate into the model different data sources and, if data are scarce, the model parameters can be updated when more data becomes available. In addition, they offer probabilistic inference given some evidence, which works in a dual direction: given evidence of an effect, inference can be conducted regarding the cause ("bottom-up" approach), and given evidence of a cause, we can infer the effect ("top-down" approach). For example, given evidence that a patient presents a series of symptoms, using a Bayesian network trained on symptom-disease data and following the "top-down" reasoning, we can estimate the probability that the patient suffers from a specific disease. These characteristics make Bayesian networks a powerful tool for RA, e.g. [69, 70].

Bayesian networks are part of the probabilistic graphical models family. They consist of nodes and edges, with nodes representing the random variables and edges the probabilistic dependencies between them. The edges of a Bayesian network are directed, modelling the causal relationship between two variables. This directionality implies a dependency relationship, making the node connected to the beginning of the edge the parent node, and the one at the end of the edge the child node. This can be further generalised by introducing the set of ancestors of a node, i.e. all nodes from

which the node at hand can be reached, as well as the set of descendants of a node; all nodes which can be reached from the node at hand. Finally, another important property of Bayesian networks is acyclicity, which, combined with the directionality of the edges, makes them directed acyclic graphs (DAGs). The lack of directional circles ensures that a node cannot belong to its set of ancestors or descendants [71].

The structure of the Bayesian network enables the computation of the joint probability distribution (JPD) through factorization, making both learning and inference simpler in terms of computational complexity. In particular, a node is independent of its non-descendants given its parents, i.e. the conditional probability distribution. Let G=(V,E) be a DAG, where V is the set of nodes and E the set of edges of the network. Then, given that X is the Bayesian network with respect to G, the JPD can be factorized as:

$$p(x) = \prod_{v \in V} p(x|x_{pa(v)})$$

where pa(v) is the set of parents of node v.

Learning and inference

The learning phase of the network comprises two components: learning of the graph topology and parameter estimation of the JPD, which in essence boils down to estimating the parameters of each conditional probability distribution (CPD), due to the factorization of the JPD. The learning process can be completely data-driven or include priors knowledge elicited from experts. In each of the two components there are two separate cases. The structure of the network can be either known or unknown, and the variables might fully or partially observable, which happens when they are latent (hidden nodes) or include missing data. All possible combinations give birth to four cases. In the first and easiest case, the structure is known and we have full observability, so the sole learning task is parameter estimation, which can be achieved through maximum likelihood estimation (MLE). Another case regards the scenario where the structure is known but is now combined with partial observability. The parameter estimates of such a problem can be calculated with iterative algorithms like expectation maximisation (EM) or gradient ascent, finding local MLEs of the parameters. In relation to the other two cases, where the structure is unknown and has to be learned through the data, the objective is to find the simplest DAG that explains the data, by applying the relevant parametric estimation depending on the observability status. More information about the learning process in Bayesian networks can be found in [72, 73]. Finally, the random variables of the network can be discrete, continuous or mixed, i.e. some variables could be discrete and others continuous. The most common and computationally tractable case is having discrete data, in which case the goal of the parametric learning is to estimate the conditional probability tables (CPT) that record the probability of the child node taking each of the discrete values for every value combination of its parents.

Inference in Bayesian networks [74] involves two approaches: predictive support and diagnostic support. In predictive support the evidence nodes (available information) are linked to the parents of the desired node, while in the case of diagnostic support, information flows opposite to the causal connections and the evidence nodes are connected to the children of the desired node [71]. In relation to the algorithmic approach, exact probabilistic inference (e.g. variable elimination) is used in simpler

problems, while approximate inference (e.g. MCMC methods) is used in complex systems.

Inference example

An example of a Bayesian network is given in Figure 24 below. The 4 variables modelled by the corresponding nodes are binary and the resulting CPTs estimated in the learning phase summarise the underlying distribution.
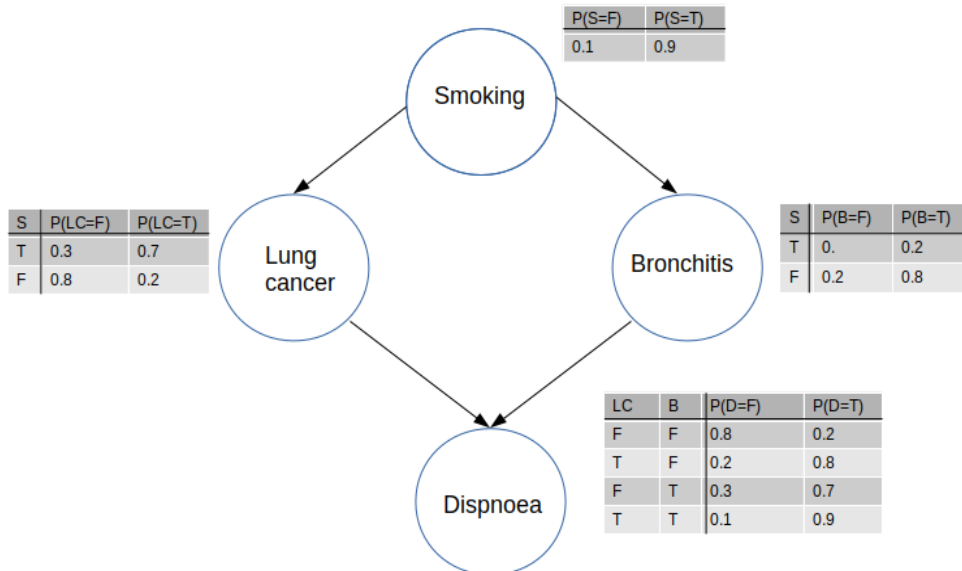

P(S=F)	P(S=T)
0.1	0.9

S	P(LC=F)	P(LC=T)
T	0.3	0.7
F	0.8	0.2

S	P(B=F)	P(B=T)
T	0.	0.2
F	0.2	0.8

LC	B	P(D=F)	P(D=T)
F	F	0.8	0.2
T	F	0.2	0.8
F	T	0.3	0.7
T	T	0.1	0.9

Figure 24: An example of trained Bayesian network.

Due to the simplified network structure, the calculation of a subset of variables given some evidence can be realised in a straightforward manner by marginalising out the rest of the variables. For instance, having information about the 'Bronchitis' node, we can calculate the probability of observing a certain state for the 'Dispnoea' node by applying the following approach:

$$p(D=T|B=T) = \frac{p(D=T,B=T)}{p(B=T)} = \frac{\sum_{LC,S\in\{F,T\}} p(D=T,B=T,LC,S)}{\sum_{LC,S,D\in\{F,T\}} p(D,B=T,LC,S)}$$

The full joint distribution presented in the numerator can be broken down into several terms using the factorization implied by the network structure, making calculations simpler:

$$\sum_{LC,S\in\{F,T\}} p(D=T,B=T,LC,S) = \sum_{LC,S\in\{F,T\}} p(D=T|B=T,LC) \cdot p(B=T|S) \cdot p(LC|S) \cdot p(S)$$

The above description of Bayesian networks reveals a series of advantages. The first one is the ability to combine data information with expert knowledge in order to derive the network topology and to estimate the model parameters. In addition, latent variables and missing data can be handled during the training phase by making use of appropriate algorithms, but more importantly, probabilistic predictions can be obtained even in data-limited environments. The variability of the incoming data is taken into account by the training and inference algorithms, enabling probabilistic reasoning induced

by observing evidence. Adding to that, uncertainty propagation is a natural effect of the conditional probabilities.[61] Finally, an equally significant feature of Bayesian networks is the natural way with which parameter estimates can be updated the moment more data become available.

Application of a Bayesian network for skin sensitisation

Even if not yet applicable to NMs, we present here as a last example the Bayesian network approach for skin sensitisation developed by Jaworska and coworkers [75, 76]. Traditionally, skin sensitisation hazards are assessed using *in vivo* animal experiments. However, legislation is increasingly being put in place to encourage the replacement of such experiments with non-animal methods and skin sensitisation has been at the centre of concerted efforts to replace animal testing in recent years. Five non-animal methods addressing mechanisms under the first three KEs of the skin sensitisation AOP have been validated by the OECD. Nevertheless, the currently adopted methods, when used in isolation, are not sufficient to fulfil regulatory requirements on the skin sensitisation potential and potency of chemicals comparable to that provided by the regulatory animal tests [77]. Therefore, the OECD initiated the development of a guideline on Defined Approaches (DAs), which is tested on different skin sensitisation approaches including the Bayesian network model.

To assess the regulatory relevance of integrated *in vitro - in silico* approaches, an important concept put forward by the OECD is the distinction between Integrated Approaches to Testing and Assessment (IATA, sometimes also called integrated testing strategies (ITS)) and DAs. A good description of the differences is given in Casati et al. [77]:

"*IATA are defined as pragmatic, science-based approaches for chemical hazard or risk assessment that rely on an integrated analysis of existing information coupled with the generation of new information using testing strategies. IATA follow an iterative approach to answer a defined question in a specific regulatory context, taking into account the acceptable level of uncertainty associated with the decision context [78]. The overall assessment process within IATA is based on WoE, which necessarily implies an expert judgment in the weighing of the different pieces of information.*

Non-animal approaches developed in the area of skin sensitisation that are based on a fixed set of information sources and fixed data interpretation procedures are designated as "defined approaches to testing and assessment" [79]. The DA designation emphasises that predictions generated by these approaches are rule-based and are not influenced by expert judgment. The fixed nature of DAs should facilitate their consideration under the OECD mutual acceptance of data (MAD), whereas IATA are designed to be flexible and adaptable to particular regional requirements or regulatory statutes."

The defined approach ITS-3 described by Jaworska et al. [75] and implemented e.g. in the SaferSkin application[3] by EwC, is designed to replace the *in vivo* Local Lymph Node Assay (LLNA) assay by using a defined approach combining results from the OECD validated DPRA [80], KeratinoSens™ [81] and h-CLAT [82] assays with predicted physico-chemical properties and predictions from the TIMES software[4] as input to the Bayesian network depicted in Figure 25. It makes a prediction for skin sensitisation potency calculated in the form of a probability distribution over 4 sensitisation classes: non-sensitizer, weak, moderate and strong sensitizer. The probability distribution is then transformed to a Bayes factor

---

[3] https://saferworldbydesign.com/saferskin/
[4] http://oasis-lmc.org/products/software/times.aspx

to remove prediction bias from the training set distribution and to give a quantitative measure to the level of uncertainty which can then be used in an objective manner to assign a confidence level to the predictions. SaferSkin is based on 207 chemicals for which physico-chemical data, *in vitro*, and *in vivo* data are available.
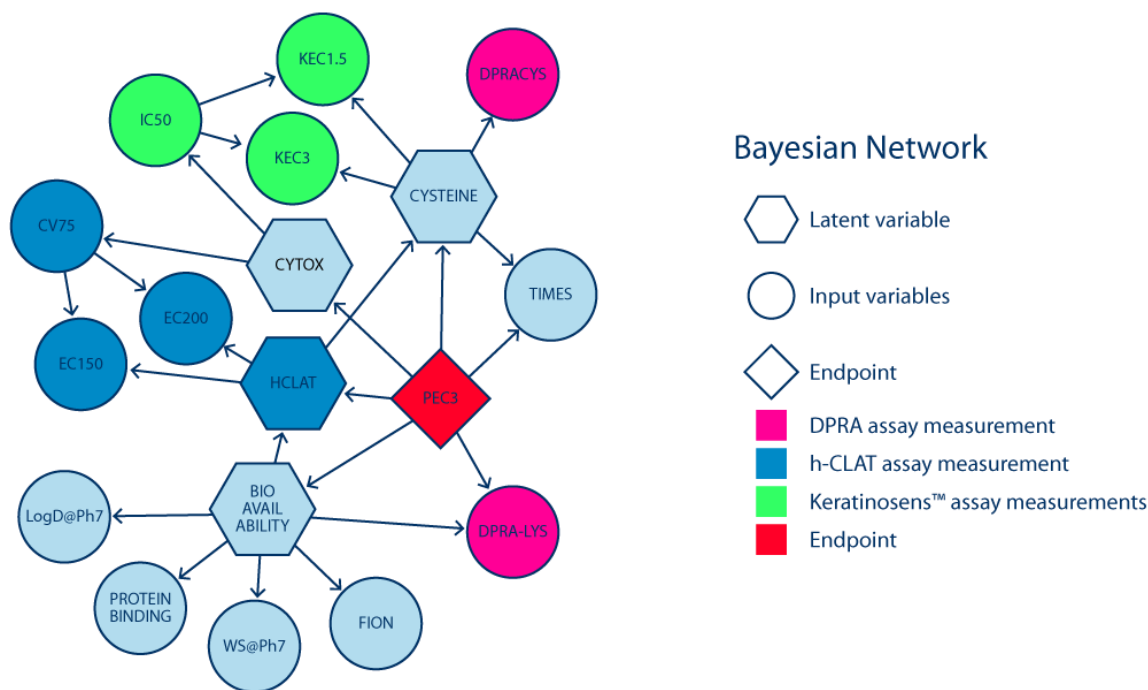
Figure 25: Layout of the Bayesian network with input nodes for the experimental results from the 3 assays, the physico-chemical properties and the TIMES result, nodes for latent variables combining related properties and the final output node predicting the results from LLNA (PEC3).

To adapt this approach to allow the usage for predicting the **skin sensitisation potency of NMs**, a number of issues have to be addressed:

1) It has to be shown that the DPRA, KeratinoSens™ and h-CLAT assay can also be applied to NMs or if this is not the case, new testing guidelines or alternative assays have to be developed. Driven by the European Commission and the EU NanoSafety Cluster, the 'Malta initiative' was established, which has the goal to develop such nano-specific guidelines with skin sensitisation being one focus. Alternative assays like Sens-IS [83] could be used to replace the current set of assays since they have broader applicability domains, i.e. they show better performance for more complex systems like mixtures, and are independent of the MIEs, which might not be direct protein binding in the case of NMs.

2) In ITS-3, biokinetics aspects influencing the exposure of the compound are approximated by the physico-chemical properties of the compounds. Corresponding properties have to be identified for NMs, which can be used as input parameters for exposure models.

3) ITS-3 predictions are depending on a QSAR model implemented in TIMES, which correlates structural features with the adversity. It has to be evaluated if such a complete *in silico* model can be created for NMs and if it improves the results from the experimental assays.

To sum up, the strength of Bayesian networks lies in the ability to train and predict in data-scarce

environments, update their parameters whenever additional data are available, and combine various data sources and formats. They can be implemented as a standalone risk analysis tool, but most importantly, their nature offers the ability to integrate responses from several tools into a predictive framework, e.g. responses from exposure and hazard models can be nodes of a Bayesian network and thus be part of an integrated analysis ending up in a risk estimation.

Future work

In relation to the NanoCommons project, we are currently investigating ways to integrate Bayesian network methodologies in NM RA, which will eventually be deployed as web services. The goal is to develop a new or reproduce an already existing Bayesian Network for NM RA using programming languages and or packages (for example R's bnlearn package) that will facilitate the creation of web services and integration into the NanoCommons knowledge infrastructure.

8. Conclusions

Overall, the approach for integration of modelling tools into the NanoCommons e-infrastructure that we presented above offers a set of favourable features. The first fulfilled aim is that our RA workflow provides a toolbox for full NM life cycle analysis from an occupational and environmental safety and health perspective. It incorporates a collection of state-of-the-art exposure models that return realistic external exposure estimations. These can be extended to further detailed, organism-specific internal exposure levels per organ/system, generated by sophisticated PBPK models, leading to a more accurate exposure estimation and hazard prediction. By taking into account the existing knowledge for the NM under investigation, as well as available data for NMs that present similar behaviour in the same environment, our approach grants reliable hazard identification and quantification. Special care is given to the minimisation of the uncertainty introduced by the reference sources used to extract Reference Points, providing a roadmap guiding the user to extract information from the most relevant source. In case multiple reference or estimated hazard values are retrieved, our workflow adopts the conservative "worst case scenario" approach. Nevertheless, the proposed optimized management of the relevant existing sources of information may facilitate the identification of cases about which more experimental data need to be generated so that sufficient information for assessing human health and environmental effects is available. Yet, by the employing dose-response models read-across algorithms and similarity assessment rules, it minimizes the need for conducting new experiments, and overcomes resource and ethical limitations. At the final step of risk characterization, but also during the previous steps of exposure and hazard evaluation, it outputs informative visualizations and reports that facilitate the outlining and communication of the results. Last but not least, this approach builds a concrete framework that bridges various diverse approaches that can be further extended and linked using data-driven methodologies, such as Bayesian Networks. For this purpose, our workflow will be executed and refined in the course of WP6 (Tools Integration) and WP9 (Dissemination and Case Studies).

In particular, this deliverable presented the breadth of modelling components that can either be used separately for particular tasks or in combination for the creation of complete RA workflows:

i) The web-based GUIDENano guidance tool which allows users to apply the most appropriate RA and risk mitigation strategy for NM-enabled products throughout their life cycle.

ii) Available tools and strategies for deriving points of departure (PODs) from dose-response data, i.e. levels of exposure that have low effect or no effect on humans or in the environment.

iii) Novel grouping/read across approaches, which combined with the repository of nanoQSAR models described in deliverable report D5.4 offers a variety of NM hazard prediction tools.

iv) The tools far for the integration of biokinetics models and especially Physiologically-Based PharmacoKinetics (PBPK) models for estimation of internal exposure of organisms to NMs.

v) Bayesian modelling approaches, which can combine both hazard and exposure information and build complete RA workflows.

Ongoing integration of these and other models and their further optimisation for NMs risk assessment will continue over the next 2.5 years.

9. References

1. Oomen, A.G., Steinhäuser, K.G., Bleeker, E.A.J., van Broekhuizen, F., Sips, A., Dekkers, S., Winjnhoven, S.W.P., Sayre, P.G., *Risk assessment frameworks for nanomaterials: Scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency*. NanoImpact, 2018. **9**: p. 1–13. doi: https://doi.org/10.1016/J.IMPACT.2017.09.001

2. Hristozov, D., Gottardo, S., Semenzin, E., Oomen, A., Bos, P., Peijnenburg, W., van Tongeren, M., Nowack, B., Hunt, N., Brunelli, A., Scott-Fordsmand, J.J., Tran, L., Marcomini, A., *Frameworks and tools for risk assessment of manufactured nanomaterials*. Environ Int, 2016. **95:** p. 36–53. doi: https://doi.org/10.1016/j.envint.2016.07.016

3. OECD (2016), "Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)", OECD Series on Testing Assessment No. 260 JT03407308, OECD Publishing, Paris, 2016. OECD Guidelines for the Testing of Chemicals, (260).

4. ECHA. "Appendix R. 6-1: Recommendations for nanomaterials applicable to the guidance on QSARs and grouping of chemicals." (2016).

5. Wijnhoven, S. W. P., et al. "Exposure to nanomaterials in consumer products." *RIVM letter report 340370001* (2009).

6. ECHA. "Guidance in a Nutshell—Chemical Safety Assessment." (2009).

7. Park M, Wijnhoven S, GUIDEnano: a web-based guidance tool for risk assessment and mitigation of nano-enabled products, TCDD 2017; (3):7

8. Cherrie JW. The effect of room size and general ventilation on the relationship between near and far-field concentrations, *Appl Occup Environ Hyg* , 1999, vol. 14 (pg. 539-46)

9. John W. Cherrie, Laura Maccalman, Wouter Fransman, Erik Tielemans, Martin Tischer, Martie Van Tongeren, Revisiting the Effect of Room Size and General Ventilation on the Relationship between Near- and Far-Field Air Concentrations, *The Annals of Occupational Hygiene*, Volume 55, Issue 9, November 2011, Pages 1006–1015, https://doi.org/10.1093/annhyg/mer092

10. Andrew D. Maynard and Anthony T. Zimmer, Development and Validation of a Simple Numerical Model for Estimating Workplace Aerosol Size Distribution Evolution Through Coagulation, Settling, and Diffusion, Aerosol Science and Technology, 37:804–817, 2003, DOI:10.1080/02786820390223963.

11. Fernández-Cruz, M.L.; Hernández-Moreno, D.; Catalán, J.; Cross, R.K.; Stockmann-Juvala, H.; Cabellos, J.; Lopes, Viviana R.; Matzke, M.; Ferraz, N.; Izquierdo, J.J.; Navas, J.M.; Park, M.; Svendsen, C.; Janer, G.. 2018 Quality evaluation of human and environmental toxicity studies performed with nanomaterials – the GUIDEnano approach. *Environmental Science: Nano*, 5 (2). 381-397. https://doi.org/10.1039/C7EN00716G

12. Park MV, Catalán J, Ferraz N, Cabellos J, Vanhauten R, Vázquez-Campos S, Janer G. Development of a systematic method to assess similarity between nanomaterials for human hazard evaluation purposes - lessons learnt. Nanotoxicology. 2018 Sep;12(7):652-676. doi: 10.1080/17435390.2018

13. NRC, US. "Risk assessment in the federal government: managing the process." *National Research Council, Washington DC* 11 (1983): 3.

14. USEPA (United States Environmental Protection Agency). "Guidelines for Carcinogen Risk Assessment, Risk Assessment Forum." (1986).

15. EPA, US. "Benchmark Dose Technical Guidance Document [External Review Draft]." *Risk

Assessment Forum, Washington, DC. 2000.

16. NRC. "Science and Decisions: Advancing Risk Assessment. National Research Council (US). Committee on Improving Risk Analysis Approaches Used by the US EPA." (2009).

17. OECD (Organisation for Economic Co-operation and Development). "Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting test guidelines 451, 452 and 453." (2012).

18. US Environmental Protection Agency. "Guidelines for carcinogen risk assessment." *Risk Assessment Forum*. Washington, Dc: US EPA, 2005.

19. US Environmental Protection Agency. "Guidelines for neurotoxicity risk assessment." *Federal Register* 63.93 (1998): 26926-26954.

20. Davis, J. Allen, Jeffrey S. Gift, and Q. Jay Zhao. "Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1. 1." *Toxicology and applied pharmacology* 254.2 (2011): 181-191.

21. European Food Safety Authority (EFSA). "Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic." *EFSA Journal* 3.10 (2005): 282.

22. Barlow, Susan, et al. "Use of the benchmark dose approach in risk assessment Guidance of the Scientific Committee." *EFSA JOURNAL* 7.6 (2009).

23. Joint, F. A. O., WHO Expert Committee on Food Additives, and World Health Organization. "Evaluation of certain food additives and contaminants: forty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives." (1995).

24. Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, K.H., More, S., Mortensen, A., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A. *Update: use of the benchmark dose approach in risk assessment*. EFSA Journal, 2017. **15**(1). doi: https://doi.org/10.2903/j.efsa.2017.4658

25. Haber, Lynne T., et al. "Benchmark dose (BMD) modeling: current practice, issues, and challenges." *Critical reviews in toxicology* 48.5 (2018): 387-415.

26. European Food Safety Authority "Use of the benchmark dose approach in risk assessment", The EFSA Journal (2009) 1150, 1-72

27. Jeliazkova, N., Chomenidis, C., Doganis, P., Fadeel, B., Grafström, R., Hardy, B., Hastings, J., Hegi, M., Jeliazkov, V., Kochev, N., Kohonen, P. et al. *The eNanoMapper database for nanomaterial safety information*. Beilstein J. Nanotechnol, 2015. 6(1): p. 1609-34. doi:10.3762/bjnano.6.165

28. Trestle Technology, LLC (2018). plumber: An API Generator for R. R package version 0.4.6. https://CRAN.R-project.org/package=plumber

29. Winkler, D. A., Mombelli, E., Pietroiusti, A., Tran, L., Worth, A., Fadeel, B., & McCall, M. J., *Applying quantitative structure–activity relationship approaches to nanotoxicology: current status and future potential*. Toxicology, 2013. **313**(1), p. 15-23. https://doi.org/10.1016/j.tox.2012.11.005

30. Gajewicz, A., Jagiello, K., Cronin, M. T. D., 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*. Environmental Science: Nano, 2017, **4**(2), p. 346-358. doi: 10.1039/C6EN00399K

31. ECHA, Read-across Assessment Framework, 2015, https://echa.europa.eu/documents/10162/13628/raaf_en.pdf.

32. Schultz, T. W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D. J., ... & Cronin, M. T. D., *A strategy for structuring and reporting a read-across prediction of toxicity*. Regulatory Toxicology and Pharmacology, 2015. **72**(3), p. 586-601. https://doi.org/10.1016/j.yrtph.2015.05.016

33. Oomen, A., Bleeker, E., Bos, P., van Broekhuizen, F., Gottardo, S., Groenewold, M., ... & Peijnenburg, W., *Grouping and read-across approaches for risk assessment of nanomaterials*. International journal of environmental research and public health, 2015. **12**(10), p. 13415-13434. https://doi.org/10.3390/ijerph121013415

34. Lamon, L., Asturiol, D., Richarz, A., Joossens, E., Graepel, R., Aschberger, K., & Worth, A., *Grouping of nanomaterials to read-across hazard endpoints: from data collection to assessment of the grouping hypothesis by application of chemoinformatic techniques*, Particle and fibre toxicology, 2018. **15**(1), 37. https://doi.org/10.1186/s12989-018-0273-1

35. ECHA, Read Across Assessment Framework, *Appendix R.6-1for nanomaterials applicable to the Guidance on QSARs and Groupingof Chemicals*. Guidance on information requirements and chemical safety assessment, 2017. https://echa.europa.eu/documents/10162/23036412/appendix_r6_nanomaterials_en.pdf

36. Gajewicz, A., Cronin, M. T., Rasulev, B., Leszczynski, J., & Puzyn, T., *Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-read-across*. Nanotechnology, 2014. **26**(1), 015701. http://dx.doi.org/10.1088/0957-4484/26/1/015701

37. Arts, J. H., Hadi, M., Irfan, M. A., Keene, A. M., Kreiling, R., Lyon, D., ... & Warheit, D., *A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping)*. Regulatory Toxicology and Pharmacology, 2015. **71**(2), S1-S27. https://doi.org/10.1016/j.yrtph.2015.03.007

38. Sayes, C. M., Smith, P. A., & Ivanov, I. V., *A framework for grouping nanoparticles based on their measurable characteristics*. International journal of nanomedicine, 2013. **8**(Suppl 1), 45. doi: 10.2147/IJN.S40521

39. Helma, C., Rautenberg, M., & Gebele, D. *Nano-lazar: read across predictions for nanoparticle toxicities with calculated and measured properties*. Frontiers in pharmacology, 2017. **8**(377). https://doi.org/10.3389/fphar.2017.00377

40. Varsou, D. D., Tsiliki, G., Nymark, P., Kohonen, P., Grafström, R., & Sarimveis, H. *toxFlow: a web-based application for read-across toxicity prediction using omics and physicochemical data*. Journal of chemical information and modeling, 2017, **58**(3), p. 543-549. https://doi.org/10.1021/acs.jcim.7b00160

41. Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... & Leszczynski, J., *Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles*. Nature nanotechnology, 2011, **6**(3), 175. doi: 10.1038/NNANO.2011.10

42. Walkey, C. D., Olsen, J. B., Song, F., Liu, R., Guo, H., Olsen, D. W. H., ... & Chan, W. C., *Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles*. ACS nano, 2014. **8**(3), p. 2439-2455. https://doi.org/10.1021/nn406018q

43. Aschberger, K., Asturiol, D., Lamon, L., Richarz, A., Gerloff, K., & Worth, A. *Grouping of multi-walled carbon nanotubes to read-across genotoxicity: A case study to evaluate the applicability of regulatory guidance*. Computational Toxicology, 2019, **9**, p. 22-35.

https://doi.org/10.1016/j.comtox.2018.10.001

44. Organization for Economic Cooperation & Development, *Case study on grouping and read-across for nanomaterials Genotoxicity of nano-TIO2*, September 2018, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2018)28&docLanguage=En.

45. Leach, A. R., and Gillet,V. J. , *An introduction to chemoinformatics*, Springer Science & Business Media, 2007.

46. Huluban, R., *Practical guide How to use and report (Q)SARs Practical Guide – How to use and report (Q)SARs*, Publications Office of the EU, 2016. doi:10.2823/81818

47. Melagraki, G. & Afantitis, A*. Enalos InSilicoNano platform: An online decision support tool for the design and virtual screening of nanoparticles*, RSC Adv, 2014. 4, p. 50713–50725. doi: 10.1039/C4RA07756C

48. Witten Ian H and Frank, Eibe and Hall, Mark A and Pal, C. J. *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016. doi:10.1016/C2009-0-19715-5

49. Jones, H.M., Rowland-Yeo, K., *Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development*. CPT Pharmacometrics Syst Pharmacol, 2013. **2**:e63. doi: https://doi.org/10.1038/psp.2013.41

50. Pilari, S., Huisinga, W., *Lumping of physiologically-based pharmacokinetic models and a mechanistic derivation of classical compartmental models*. J Pharmacokinet Pharmacodyn, 2010. **37**(4): p. 365–405. doi: https://doi.org/10.1007/s10928-010-9165-1

51. Nestorov, I.A., Aarons, L.J., Arundel, P.A., Rowland, M., *Lumping of whole-body physiologically based pharmacokinetic models*. J Pharmacokinet Biopharm, 1998. **26**(1): p. 21–46. doi: https://doi.org/10.1023/A:1023272707390

52. Sturm, R.A., *Computer Model for the Clearance of Insoluble Particles from the Tracheobronchial Tree of the Human Lung*. Comput Biol Med, 2007. **37**(5): p. 680–690. doi: https://doi.org/10.1016/j.compbiomed.2006.06.004

53. Brown, R.P., Delp, M.D., Lindstedt , S.L., Rhomberg, L.R., Beliles, R.P., *Physiological parameter values for physiologically based pharmacokinetic models*. Toxicol Ind Health, 1997. **13**(4): p. 407–84. doi: https://doi.org/10.1177/074823379701300401

54. Edginton, A.N., Theil, F.P., Schmitt, W., Willmann, S., *Whole body physiologically-based pharmacokinetic models: their use in clinical drug development*. Expert Opin Drug Metab Toxicol, 2008. **4**(9): p. 1143-1152. doi: https://doi.org/10.1517/17425255.4.9.1143

55. Zhuang, X., Lu C., *PBPK modeling and simulation in drug research and development*. Acta Pharmaceutica Sinica B, 2016. **6**(5): p. 430-440. doi: https://doi.org/10.1016/j.apsb.2016.04.004

56. Nestorov, I., *Whole body pharmacokinetic models*. Clin Pharmacokinet, 2003. **42**(10): p. 883-908. https://doi.org/10.2165/00003088-200342100-00002

57. Li, M., Al-Jamal, K.T., Kostarelos, K., Reineke, J., *Physiologically Based Pharmacokinetic Modeling of Nanoparticles*. ACS Nano, 2010. **4**(11): p. 6303–6317. doi: https://doi.org/10.1021/nn1018818

58. Yuan, D., He, H., Wu, Y., Fan, J., Cao, Y., *Physiologically Based Pharmacokinetic Modeling of Nanoparticles*. J Pharm Sci, 2019. **108**(1): p. 58–72. doi: https://doi.org/10.1016/j.xphs.2018.10.037

59. Li, D., Johanson, G., Emond, C., Carlander, U., Philbert, M., Jolliet, O., *Physiologically based pharmacokinetic modeling of polyethylene glycol-coated polyacrylamide nanoparticles in rats*.

Nanotoxicology, 2014. **8**: p. 128–137. doi: https://doi.org/10.3109/17435390.2013.863406

60. Carlander, U., Li, D., Jolliet, O., Emond, C., Johanson, G., *Toward a general physiologically-based pharmacokinetic model for intravenously injected nanoparticles*. Int J Nanomed, 2016. **11**: p. 625–640. doi: https://doi.org/10.2147/IJN.S94370

61. Wiesner, M.R., Bottero, J-Y., *A risk forecasting process for nanostructured materials, and nanomanufacturing*. Comptes Rendus Physique, 2011. **12**(7): p. 659-668. doi: https://doi.org/10.1016/j.crhy.2011.06.008

62. Marvin, H.J.P., Bouzembrak, Y., Janssen, E.M., van der Zande, M., Murphy, F., Sheehan, B., Mullins, M., Bouwmeester, H., *Application of Bayesian networks for hazard ranking of nanomaterials to support human health risk assessment*. Nanotoxicology, 2017. **11**(1): p. 123-133. doi: https://doi.org/10.1080/17435390.2016.1278481

63. Money, E.S., Reckhow, K.H., Wiesner, M.R., *The Use of Bayesian Networks for Nanoparticle Risk Forecasting: Model Formulation and Baseline Evaluation*. Sci Total Environ, 2012. **426**: p. 436–445. doi: https://doi.org/10.1016/j.scitotenv.2012.03.064

64. Money, E.S., Barton, L.E., Dawson, J., Reckhow, K.H., Wiesner, M.R., *Validation and sensitivity of the FINE Bayesian network for forecasting aquatic exposure to nano-silver*. Sci Total Environ, 2014. **473–474**: p. 685–691. doi: https://doi.org/10.1016/j.scitotenv.2013.12.100

65. Murphy, F., Sheehan, B., Mullins, M., Bouwmeester, H., Marvin, H.J.P., Bouzembrak, Y., Costa, A.L., Das, R., Stone, V., Tofail, S.A.M., *A tractable method for measuring nanomaterial risk using Bayesian networks*. Nanoscale Res Lett, 2016. **11**(1): p. 503. doi: https://doi.org/10.1186/s11671-016-1724-y

66. Gerloff, K., Landesmann, B., Worth, A., Munn, S., Palosaari, T., Whelan, M., *The adverse outcome pathway approach in nanotoxicology*. Comput Toxicol, 2016. **1**: p. 3-11. doi: https://doi.org/10.1016/j.comtox.2016.07.001

67. Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., *Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment*. Environ Toxicol Chem, 2010. **29**(3): p. 730–741. doi: https://doi.org/10.1002/etc.34

68. Jeong, J., Song, T., Chatterjee, N., Choi, I., Cha, Y.K., Choi, J., *Developing adverse outcome pathways on silver nanoparticle-induced reproductive toxicity via oxidative stress in the nematode Caenorhabditis elegans using a Bayesian network model*. Nanotoxicology, 2018. **12**(10): p. 1182-1197, doi: 10.1080/17435390.2018.1529835

69. Beaudequin, D., Harden, F., Roiko, A., Stratton, H., Lemckert, C., Mengersen, K., *Beyond QMRA: modelling microbial health risk as a complex system using Bayesian networks*. Environ Int, 2015. **80**: p. 8-18. doi: https://doi.org/10.1016/j.envint.2015.03.013

70. Carriger, J.F., Martin, T.M., Barron, M.G., *A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors*. Aquat Toxicol, 2016. **180**: p. 11–24. doi: https://doi.org/10.1016/j.aquatox.2016.09.006

71. Ben-Gal, I., *Bayesian networks*. In: Encyclopedia of Statistics in Quality and Reliability I, 2008. doi: https://doi.org/10.1002/9780470061572.eqr089

72. Cheng, J., Greiner, R., Kelly, J., Bell, D., Liu, W., *Learning Bayesian networks from data: An information-theory based approach*. Artificial Intelligence, 2002. **137** (1-2): p. 43-90. doi: https://doi.org/10.1016/S0004-3702(02)00191-1

73. Heckerman, D., *A Tutorial on Learning with Bayesian Networks*. In: Holmes D.E., Jain L.C. (eds)

Innovations in Bayesian Networks. Studies in Computational Intelligence, vol 156. 2008: Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-540-85066-3_3

74. Huang, C., Darwiche, A., *Inference in belief networks: A procedural guide*. Int J Approx Reason, 1996. **15**(3): p. 225-263. doi: https://doi.org/10.1016/S0888-613X(96)00069-2

75. Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T, Miyazawa, M., *Bayesian Integrated Testing Strategy (ITS) for Skin Sensitization Potency Assessment: A Decision Support System for Quantitative Weight of Evidence and Adaptive Testing Strategy.* Arch Toxicol, 2015. **89** (12): p. 2355–83. doi: https://doi.org/10.1007/s00204-015-1634-2

76. Pirone, J.R., Smith, M., Kleinstreuer, N.C., Burns, T.A., Strickland, J., Dancik, Y., Morris, R., Rinckel, L.A., Casey, W., Jaworska, J.S., *Open Source Software Implementation of an Integrated Testing Strategy for Skin Sensitization Potency Based on a Bayesian Network*. ALTEX, 2014. **31**(3): p. 336–40, doi: https://doi.org/10.14573/altex.1310151

77. Casati, S., Aschberger, K., Barroso, J., Casey, W., Delgado, I., Kim, T. S., Kleinstreuer, N., Kojima, H., Lee, J. K., Lowit, A., Park, H.K., Régimbald-Krnel, M.J., Strickland, J., Whelan, M., Yang, Y., Zuang, V., *Standardisation of Defined Approaches for Skin Sensitisation Testing to Support Regulatory Use and International Adoption: Position of the International Cooperation on Alternative Test Methods*. Arch Toxicol, 2018. **92**(2): p. 611–17, doi: https://dx.doi.org/10.1007%2Fs00204-017-2097-4

78. OECD (2017b) Guidance document on the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA). Series on testing & assessment no. 260. http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

79. OECD (2016a) Guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitisation. Series on testing & assessment no. 256. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29&doclanguage=en.

80. OECD (2015a) Guidelines for the testing of chemicals test no. 442C. In chemico skin sensitisation: direct peptide reactivity assay (DPRA). http://www.oecdilibrary.org/environment/test-no-442c-in-chemico-skin-sensitisation_9789264229709-en.

81. OECD (2015b) Guidelines for the testing of chemicals test no. 442D. In vitro skin sensitisation: ARE-Nrf2 Luciferase test method. http://www.oecdilibrary.org/environment/test-no-442d-invitro-skin-sensitisation_9789264229822-en.

82. OECD (2017a) OECD test guideline no. 442E: in vitro skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation. http://www.oecdilibrary.org/environment/test-no-442e-in-vitro-skin-sensitisation_9789264264359-en.

83. Cottrez, F., Boitel, E., Ourlin, J-C., Peiffer, J-L., Fabre, I., Henaoui, I-S., Mari, B., Vallauri A., Paquet, A., Barbry, P., Auriault, C., Aebye, P., Grouxa, H., *SENS-IS, a 3D Reconstituted Epidermis Based Model for Quantifying Chemical Sensitization Potency: Reproducibility and Predictivity Results from an Inter-Laboratory Study.* Toxicol In Vitro, 2016. **32**: p. 248–60, doi: https://doi.org/10.1016/j.tiv.2016.01.007

Annexes

Annex 1: Data provided to a GUIDEnano case study

For the purpose of showing the data that a user is asked to enter in GUIDEnano we use a "simple" case study extracted from a published paper looking at the release of nano-TiO_2 during the fabrication of a paint. The worker is pouring a bag of nano-TiO_2 powder into a liquid paint matrix in a factory hall. DOI: 10.1039/C4EM00532EScenario. In all the following tables the data entry for the activity, nano(materials), compartment and Hazard exposure modules are listed.

Activity			
General Info	Name	Pouring 25 kg RD3.	
	Setting scale	Large industry	
	Handling type	Manual	
	Applied energy level	Medium	
	Life cycle phase	Production	
	Concurrent locations	1	
Input, output and release	Activity input	TiO_2 NM poured during the activity, Material: TiO2 RD3, Total amount: 26, Unit: kg, Ref.: yes, Rate: 26 kg/min	
	Activity output	TiO_2 contained in the formulated paint, Material: TiO_2 RD3, Relative to: Input/ TiO_2 NM poured during the activity, Relative amount: 99.9%, Total amount: 25.9985, Unit: kg, Ref.: no.	
	Activity release	Emitted particles into the room (indoor), Material: TiO_2 RD3, Relative to: Input | TiO_2 NM poured during the activity, Relative release: 0.0005512%, RMM: no, Total release: 0,00143312, Unit: kg, Ref.: no, Rate/location: 143.3 mg/min.	
Duration	Activity repetition	Number of times this activity is applied for the same batch: 10.	
	Operational repetition	operational time needed to complete this activity: 1 min	

	Time span	Total time span of all activity cycles together: 20 min.
(Nano)materials flow	input	Input: TiO_2 NMs that are poured during the activity.
	output	Output(s): TiO_2 contained in the formulated paint.
	release	Release(s): Release | Emitted particles into the room (indoor), Into compartment | zone: Factory hall | NF (LCLZ).

(Nano)materials		
Identification	name	TiO_2 RD3
	description	Pigment GmbH, Pori, Finland, CAS 13463-67-7
	origin	Engineered
	Source/supplier	Sachtleben Pigment GmbH, Pori, Finland, CAS 13463-67-7
Shape and Size	shape	Spherical
	Size distribution available	yes
	Method used	other
	Size type	Primary size
	Metric of size distribution	mass
	Mean size	220
	Standard deviation	22
Physical properties	Physical state	Solid
	Size category this solid materials present	Ultrafine powder (100 nm – 1 μm); nanoscale particles (1 nm-100 nm)
	Rigidity	rigid
	Dustiness [mg/kg]:	5.3

Surface properties	Layout and charge	Chemical compound: TiO3 RD3, Role: core, Al_2O_3, and: ZrO2
Other properties	function	Pigment, UV filters
	Chemical info	Are all constituents, impurities and contaminants added and identified? Yes. Purity in % :100
Constituents	Mass density	$4\ g/cm^3$
	Select constituents	Category: chemical, name/identifier: TiO2 RD3 (CAS No.13463-67-7), phase: solid, role of constituent: core, conc.:93, unit: %, mass perc.: 93%
	Select constituents	Category: chemical, name/identifier: Al2O3, phase: solid, role of constituent: coating, conc.: 3.5, unit: %, mass perc.: 3.5%
	Select constituents	Category: chemical, name/identifier: ZrO_2, phase: solid, role of constituent: coating, conc.: 3.5, unit: %,mass perc.: 3.5%.

Compartment		
	Select compartment	Type: indoor air, Name: Factory hall
		Type outdoor air, Name: factory hall
Factory hall		
General	Name:	Factory hall
	With of the room	20 m
	Length of the room	30 m
	Volume of the room	1500 m3

Zones	Select zone:	Zone description: NF (LCLZ), Number: 1, Medium: air, Size: 8, Unit: m^3, Total dimension: 8 m^3.
	Select zone:	Zone description: Floor, Number: 1, Medium: solid, Size: 600, Unit: m^2, Total dimension: 600 m^2.
	Select zone:	Zone description: Rest of the room (FF), Number: 1, Medium: air, Size: 1492, Unit: m^3, Total dimension: 1492 m3.
Zone: NF (LCLZ)	Properties:	Temperature: 25.0 °C.
	Contact zones:	In contact with: Floor, Orientation: below, Separated: virtually. In contact with: Rest of the Room (FF), Orientation: around, Separated: virtually.
	Exposed	Select or add a new exposed human population or eco species: Workers | exposure NF (LCLZ).
Zone: Floor	Properties:	Temperature: 25.0 °C.
	Contact zones:	In contact with: NF (LCLZ), Orientation: above, Separated: virtually. In contact with: Rest of the Room (FF), Orientation: above, Separated: virtually.
	Exposed	Select or add a new exposed human population or eco species: Workers | exposure Floor.
Zone: Rest of the Room (FF)	Properties:	Temperature: 25.0 °C, Pressure: 1 atm., Mechanical ventilation: Yes, Air exchanges per hour [/h]: 5.
	Contact zones:	In contact with: NF (LCLZ), Orientation: within, Separated: virtually. In contact with: Floor, Orientation: below, Separated: virtually. In contact with: Outdoor air (outside of the factory hall) | outdoor air, Orientation: around, Separated: physically.
	Exposed	Select or add a new exposed human population or eco species: Workers | exposure Rest of the room (FF).

Exposure/Hazard assessment		
	Select human populations:	Population name: Workers, Group: workers

General	Population name:	Workers
	Population category	Workers
Exposure paths	Select indirect through zones:	Exposure zone(s): factory hall | NF (LCLZ), Route(s): inhalation, Exposure relevant material: TiO2 RD3.
	Select indirect through zones:	Exposure zone(s): factory hall | Rest of the room (FF), Route(s): inhalation, Exposure relevant material: TiO2 RD3.
	Select indirect through zones:	Exposure zone(s): factory hall | Floor, Route(s): dermal, Exposure relevant material: TiO2 RD3.

Annex 2: Checklist for NanoCommons workflow for Risk Assessment

Information Gathering

R1.　The user provides the information about the nanomaterial (NM) of interest

　　R1.1.　The user provides information about the chemical composition and purity of the NM

　　R1.2.　The user provides information about the impurities contained in the NM and their quantities

　　R1.3.　The user provides the NM primary size distribution and the respective measurement method

　　R1.4.　The user provides the aggregate/agglomerate size distribution

　　R1.5.　The user provides the NM physical properties:
- phase
- rigidity
- density
- crystalline form
- melting point
- boiling point

　　R1.6.　The user provides NM Reactivity information

　　R1.7.　The user provides information about the NM Shape

　　R1.8.　The user provides information about NM Surface properties:
- charge
- zeta potential
- solubility
- hydrophobicity/hydrophilicity

　　R1.9.　The user provides information about the NM Function

　　R1.10.　The user provides information about the NM Classification (Hazard Statement)

　　R1.11.　More than one NM can be included in each exposure scenario

R2.　The user provides information about the Activity during which the exposure to NM occurs:

　　R2.1.　The user provides general information about the Activity related to the NM exposure:
- Scale
- Handling
- Applied energy level
- NM life cycle phase
- Concurrent locations

　　R2.2.　The user provides specific information about the Activity related to the NM exposure:
- Input
- Output
- Release
- Operational time
- Repetition

　　R2.3.　The user provides information about the NM flow (between compartments) during the exposure

　　R2.4.　The user provides information about the Local controls, if available

　　R2.5.　More than one activity can be included in each exposure scenario

R3.   The user provides General information about one or more compartments (indoor and environmental): Dimensions, surface area, volume(s)

R4.   The user provides information about the separate Zones:
- General info: ventilation, air properties
- Composition
- Contact zones
- Emmision(s) [see activity description]
- Exposed Species: Human, animal.

R5.   The user provides information about the Distance from NM source

R6.   The user provides information about the Exposure route:
- Inhalation
- Oral
- Dermal.

Exposure Assessment

R7.   The user chooses one or more of the following options:
R7.1.   The user enters local measurements of the NM concentration
R7.2.   The user enters an estimation of the NM concentration
R7.3.   The user chooses one or more models to use for exposure estimation:
- Zone derived estimate
- ART
- Marina Library estimate.

R8.   The user chooses whether a PBPK model will be used to derive the tissue/organ-specific mass profiles of NMs.
R8.1.   If a relevant PBPK model, specific to the NM and administration route, has been uploaded on Jaqpot:
R8.1.1.   If the model refers to different species, the user performs extrapolation e.g. by allometric scaling.
R8.1.2.   The user provides the External exposure estimate as dose in the corresponding input compartment.
R8.1.3.   The user provides the model parameters (e.g. physiological covariates, if applicable).
R8.2.   If no relevant PBPK model is available, the user should request the upload of a corresponding model.

Hazard Assessment

R9.   The user provides reference information about the FSLV(s) depending on the aim of the Hazard assessment:
- DNEL
- NOAEL/LOAEL
- OEL
- PNEC

R10.   If there are no available Reference Points, the user provides the following information to the PROAST tool:

- Dose - Response experimental data
- The name of the dose variable
- The name of the response variable(s)
- The BMR percentage (default 5%)
- Number of bootstraps (default 10,000).

R11.   If there are no available Dose - Response experimental data:

    R11.1.   If there are available characterization data for more than 20 samples, the user is prompted to prefer to use a nano-QSAR model:

- The user provides the characterization data.

    R11.2.   Or a Read-across methodology can be implemented:

- The user provides NM grouping data.

    R11.3.   If there are limited available data, then the user is prompted to use a Read-Across across methodology by providing a minimal set of descriptors, for example physicochemical properties, theoretical descriptors etc.

- The user enters Other model parameters (depending on the model: grouping or analog approach).

    R11.4.   If there are no other available data resources the user is prompted to perform Toxicological Studies similarity assessment in order to choose candidate toxicology studies to extract quantitative Hazard information:

- The user chooses from a list of available studies that are characterized as "Accepted" by the NM similarity, quality and relevance of each study assessment module.

Risk Assessment

R12.   The user may select among the available measured or estimated Exposure levels.

R13.   The user may select among the available reference, estimated or approximated Hazard values.

R14.   The user may select available protective controls or exposure modification factors to evaluate their effect on the RCR for each assessed exposure scenario.

Annex 3: Catalogue of nano-PBPK models found in the literature

Nanoparticle	Species	Administration route	Reference
Quantum Dot 705 (QD705)	Mouse	Intravenous	http://doi.org/10.1021/es800254a
[99m]Technetium-labelled carbon nanoparticles (Technegas)	Male human	Pulmonary	http://doi.org/10.3109/08958370902748542
Silver nanoparticles	Rat	Intravenous	http://doi.org/10.1016/j.biomaterials.2010.07.045
poly(lactic-co-glycolic) acid (PLGA) nanoparticles	Mouse	Intravenous	http://doi.org/10.2147/IJN.S23758
gold/dendrimer composite nanodevices (CNDs)	Mouse	Intravenous	http://doi.org/10.1007/s11095-012-0784-7
Silver, silver alloy and ionic silver	Rat	Intravenous	http://doi.org/10.2147/IJN.S46624
[192]Ir radio-labeled iridium particles and ultrafine carbon particles	Rat	Pulmonary	http://doi.org/10.1115/1.4025333
Polyacrylamide (PAA)	Rat	Intravenous	http://doi.org/10.3109/17435390.2013.863406
Gold nanoparticles	Mouse	Intravenous	https://doi.org/10.3109/17435390.2015.1027314
Silver nanoparticles	Rat	Intragastric	http://doi.org/10.1134/S1995078015020081
Titanium dioxide (TiO$_2$) nanoparticles	Mouse, Rat	Intravenous	http://doi.org/10.3109/17435390.2014.940404

Zinc oxide (ZnO) and Zinc Nitrate (Zn(NO$_3$)$_2$) nanoparticles	Mouse	Intravenous	http://doi.org/10.2147/IJN.S86785
Cadmium telluride/cadmium sulphide quantum dots (CdTe/CDS Qds)	Mouse	Intravenous	http://doi.org/10.1021/acs.nanolett.5b03854
Silver and Carbon black (CB) nanoparticles	Mouse	Pulmonary	http://doi.org/10.1371/journal.pone.0080917
PAA, Gold, TiO$_2$	Rat	Intravenous	http://doi.org/10.2147/IJN.S94370
QD705	Mouse	Intravenous	http://doi.org/10.1021/nl803481q
TiO$_2$	Rat	Intravenous	http://doi.org/10.1371/journal.pone.0124490
Amphiphilic block copolymers poly (ethylene glycol) (PEG) and poly (ε-caprolactone) (PCL) bearing pendant cyclic ketals	Mouse	Intravenous	http://doi.org/10.1002/psp4.13
Nanocrystals of SNX-2112	Rat	Intravenous	http://doi.org/10.2147/IJN.S79734
[192]Ir radio-labelled iridium and silver nanoparticles	Rat	Pulmonary, Intravenous	https://doi.org/10.1016/j.yrtph.2015.06.019
Gold	Mouse, rat, pig	Intravenous	http://doi.org/10.2217/nnm.15.177
PAA	Rat	Intravenous	https://doi.org/10.1016/j.taap.2010.11.017
(192)Ir radio-labelled iridium and silver nanoparticles	Rat	Pulmonary	https://doi.org/10.1088/1742-6596/151/1/012028

Gold	Rat	Pulmonary	https://doi.org/10.1088/1742-6596/151/1/012029
Molecular imaging nanoparticles (MINs)	Mouse	Injection	https://doi.org/10.1089/oli.2009.0216
Cerium dioxide (CeO$_2$)	Rat	Pulmonary	http://doi.org/10.1186/s12989-016-0156-2
CeO$_2$	Rat	Pulmonary, Intravenous, oral or its instillation	http://doi.org/10.2147/IJN.S157210
Superparamagnetic iron oxide nanoparticles (SPIONs)	Mouse	Intravenous	http://doi.org/10.1515/ejnm-2017-0001