# 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 5.1

| | |
|---|---|
| **Deliverable** | D5.1 Initial Knowledge Infrastructure Functionalities and Services Implemented |
| **Work Package** | WP5: Analysis and Modelling tools |
| **Delivery date** | M12 - 31 December 2018 |
| **Lead Beneficiary** | Douglas Connect (DC) and University of Birmingham (UoB) |
| **Nature of Deliverable** | Demonstrator |
| **Dissemination Level** | Public (PU) |

| | |
|---|---|
| **Submitted by** | Thomas Exner and Lucian Farcal (DC), Anastasios diamantis and Iseult Lynch (UoB), Vladimir Lobaskin (NUID ), Philip Doganis and Haralambos Sarimveis (NTUA), Antreas titis (NovaM), Dieter Maier (BIOMAX), Egon Willighagen (UM) |
| **Revised by** | Anastasios Papadiamantis (UoB), Lucian Farcal (DC) |
| **Approved by** | Iseult Lynch (UoB) |

# Table of contents

# Summary

The scope of Work Package 5 (WP5) is the integration of the state-of-the-art tools for data management, mining, handling, analysis and predictive modelling to facilitate their joint provision as tools and services for Transnational Access (TA). To achieve this, a linked data approach is being utilised to exploit, extract, and integrate knowledge from all available information types (raw and processed experimental and modelling data, and metadata) to be captured in the NanoCommons Knowledge Base (KB). These tools, once linked and interoperable via the NanoCommons platform, will be made available to the nanosafety community either as standalone or through the NanoCommons TA open calls that will offer funded access to the tools and services offered by NanoCommons to develop tailored solutions for individual user or project's data management needs. The TA calls will allow successful applicants to take advantage of the NanoCommons tools and services free-of-charge, including help to make their existing data FAIR (Findable, Accessible, Interoperable, and Reusable), to implement data management, mining, handling, analysis, visualisation and modelling workflows into their experimental design and data collection processes and to facilitate direct integration of resulting datasets into relevant databases (experimental and *in silico*) with all of the required metadata and semantic annotation. NanoCommons TA Users are thus becoming part of a wider attempt at cross-field and cross-data harmonisation, leading to increased data quality and thus better predictive models for nanosafety, thereby resulting in ground breaking insights and novel materials.

This advance in nanosafety and safe-by-design (SdD) NMs will also be facilitated through the further development of the existing tools and the design and creation of new tools to fill in any identified research gaps or lack of services. This deliverable (D5.1) presents the tools and respective services implemented during Year 1 within NanoCommons that are offered during the first TA call. Additional services will continue to be added to the NanoCommons portfolio and refinements to current services will also be made based on experience gained during the first round of TA provision.

# Abbreviations

AOP: Adverse Outcome Pathway

API: Application Programming Interface

CAS: Chemical Abstracts Service

CG: Coarse Grained

CIS: Configuration Interaction Singles

COSMO: Conductor-like Screening Model

DFT: Density Functional Theory

DFTB: Density Functional Tight-Binding

DNA: Deoxyribo Nucleic Acid

DW: Data Warehouse

EC: European Commission

ELN: Electronic lab notebook

ESPResSo: Extensible Simulation Package for Research on Soft Matter

EZ: Embryonic Zebrafish (assay)

FAIR: Findable, Accessible, Interoperable, and Reusable

FP7: Framework Program 7

GO: Gene Ontology

GROMACS: GROningen MAchine for Chemical Simulations (computational software)

GSEA: Gene-Set-Enrichment Analysis

H2020: Horizon 2020

HCNP: Hydrophobic Charged Nanoparticles

HF: Hartree-Fock

iTasser: Iterative Threading Assembly Refinement

KB: KnowledgeBase

KI: Knowledge Infrastructure

LDM: Liquid Drop Model

MD: Molecular Dynamics

MI: Mutual Information

NM: Nanomaterial

NBI: Nanomaterial-Biological Interactions

NIKC: Nanoinformatics Knowledge Commons

NWChem: North West Chemistry (computational software)

OECD: Organisation for Economic Cooperation and Development

PM6 or PM7: Parametrization Method 6 or 7

QSARs: Quantitative Structure Activity Relationships

RT: Request Tracker

SdD: Safe-by-design

SNT-MT: SmartNanoTox Modelling Tools

TA: Transnational Access

TDDFT: Time-Dependent Density Functional Theory

vdW: van der Waals

VSN: Variance Stabilisation by Normalisation

WP: Work Package

# 1. Introduction

Nanosafety research is being rapidly transformed into a highly competitive and data intensive field. In parallel, translational research, the introduction of nanoinformatics and the shift towards big data have introduced the need for more complex data capturing, storage, handling and analysis tools and processes. At the same time, the advancement of machine learning and modelling and analysis techniques have raised the demand for large volumes of high quality findable, accessible, interoperable, and reusable (FAIR)[1] data [1]. Currently, much of the nanosafety data produced is fragmented, scattered and inaccessible and does not enable the development and application of modelling and read-across approaches, which are currently absent for nanomaterials. As a result, the cost of nanosafety research and the cost and time required for regulation and market approval significantly increases, stifling data-driven innovation.

To address these gaps, NanoCommons has developed a nanosafety knowledge infrastructure (KI) to organise, harmonise and make nanosafety data FAIR, while at the same time linking it with the necessary tools for high quality data management, mining, handling, analysis, visualisation and modelling. The aim of NanoCommons is to deliver a sustainable and openly accessible nanoinformatics framework for assessment of the risks of nanomaterials (NMs), their products and their formulations. The knowledge infrastructure aims to cover the whole data lifecycle, from experimental design, selection of protocols and methods, to data capture templates such that data management workflows are integrated into experimental workflows, through to data enrichment, data mining and application of nanoinformatics approaches to the datasets. This integrated KI is made accessible to all stakeholders (academia, industry, regulators, etc.) enabling them to address their data and research needs.

Work Package 5 (WP5) aims to integrate the current state of the art tools for data management, data mining and data analysis, utilising a linked data approach that will exploit and integrate knowledge from all available information sources (raw experimental and modelling data, and metadata) and enable their linkage into the KI. These tools, once available and interoperable via the NanoCommons platform, will be offered to the nanosafety community via open calls for funded access. NanoCommons is not looking to "reinvent the wheel" and develop new services and tools, but rather to use, adapt and further improve already existing tools where possible, especially with respect to accessibility and user-friendliness to fit the needs of the experimental nanosafety community who are currently the largest generators and potential providers of data for nanoinformatics and modelling. For example, existing data for physics and chemistry-based materials modelling procedures will be adapted to calculate relevant NMs descriptors and complement the data sets where information gaps are identified. Existing data handling and analysis tools will be further developed, extended and integrated throughout the project, taking into account existing knowledge from chemicals, the additional needs of the nanosafety community due to the larger and more diverse data sources and NM structures. Extracted knowledge will then be organised in formats suitable for predictive modelling. The tools used and developed within WP5 will be implemented, whenever possible, based on interoperable, standards-compliant modular web services maximising cross-talk and interaction between different/diverse sources of data.

---

[1] https://www.force11.org/group/fairgroup/fairprinciples

The integrated tools provided by NanoCommons aim to address the following issues:
- extracting knowledge from raw experimental data (such as microscopic images or spectral data);
- preprocessing the data before they are sent to modelling services (normalisation, missing data handling, selection of important variables, dimensionality reduction);
- generating theoretical descriptors (such as structural descriptors or quantum mechanical descriptors);
- analysing big omics or "corona" data in terms of identifying the biological mechanisms and pathways associated with toxicity and other adverse effects and producing aggregated biologically enriched descriptors;
- harmonising and integrating diverse data and metadata originating from heterogeneous resources, so that homogeneous datasets suitable for modelling are produced;
- semantically retrieving ontology annotated data from the public sources including the project data warehouse; and
- continuously supporting the NanoCommons Users by providing the optimal training and use of the integrated tools.

As a result, the services so far offered through NanoCommons, and presented in detail below, include:

- Data and knowledge bases:
  - NanoMILE physicochemical characterisation library and *in vitro* toxicity data
  - NanoFASE nanomaterials physicochemical characterisation library
  - The Nanomaterial-Biological Interactions (NBI) Knowledgebase
  - ACEnano data warehouse
  - eNanoMapper database and project data warehouses based on this data management system (NanoWiki, NANoREG data)
  - WikiPathways SPARQL endpoint for biological pathway knowledge
  - AOPWiki SPARQL endpoint (under development) for adverse outcome pathways
- Data management tools and workflows:
  - Electronic notebooks
  - Data and metadata curation tools
  - Data and metadata semantic annotation tools
  - Protocols and methods repository
  - Semantic annotation of data and metadata
  - Edelweiss data sharing platform
- Development and implementation of experimental workflows:
  - ENM physicochemical characterisation
  - Toxicology and Ecotoxicology
  - Biomolecule corona analysis (protein, lipid, carbohydrate, small molecule)
- Risk assessment and mitigation strategies:
  - Guidenano tool
- Omics data analysis:
  - Evaluation of data quality
  - Normalisation

- ○ Differential expression analysis
  - ○ Functional enrichment analysis
  - ○ Network reconstruction
- *In silico* modelling:
  - ○ Statistical modelling, data-driven NM property and functionality prediction, Quantitative Structure Activity Relationships (QSARs)
  - ○ Materials modelling, physics-based NM property prediction
- Image analysis tools:
  - ○ NanoImage (NTUA)
  - ○ NanoXtract (NM)
- Dedicated Helpdesk

As a result, NanoCommons has the unique potential to deliver a step-changing impact for the emerging nanoinformatics in nanosafety community. It will remove barriers from nanosafety data sharing, and ensure data accessibility for nanomaterials regulators and industry by revolutionising data capture, management, analysis and sharing, thus creating a global community that will be able to address current and future scientific and regulatory issues.

## 2. NanoCommons Services Catalogue

The central entry point to the NanoCommons KB is the NanoCommons Service Catalogue, directly accessible from the NanoCommons website. It is based on the technology developed in the e-infrastructure project OpenRiskNet[2], which is offered as a free service to other European, national and international projects, which has been specifically adapted to the needs of NanoCommons. The catalogue supports the users in finding information on TA offerings and the corresponding tools and services used within these, as well as how self-learners can access and use the infrastructure functionalities and individual services through an integrated information and support system.

The catalogue provides a detailed description (Table 1) of the available services, integrated or in the process of being integrated, and provides direct links to the service environment and/or to the Application Programming interface (API) definition and also to all related support resources provided by OpenRiskNet or the original service provider. The services can be filtered by application area, service type and user role. Moreover, the catalogue features tags describing the status of the integration (complete, underway, planned) into the NanoCommons KB and specific TA offerings.

**Table 1**. The structure of service description

| | |
|---|---|
| **Service identification** | <ul><li>Date of Creation</li><li>Date of updates</li><li>Service name</li><li>URL</li><li>API URL</li><li>API Type</li><li>Provider name</li><li>Provider contact</li><li>Provider organisation</li><li>Origin of the service (e.g. provided by OpenRiskNet or third parties)</li><li>Category of service</li><li>Service type</li><li>Integration status</li><li>Implementation status</li><li>Licence type</li><li>Licence</li><li>Information if login is required</li></ul> |
| **Service description** | <ul><li>Short description</li><li>Applicability domain</li><li>Topic</li><li>Biological area covered</li><li>User type (developers or end-users)</li><li>Targeted industry</li><li>Targeted users</li><li>Relevant case studies</li></ul> |
| **Training and user support** | <ul><li>User support service</li><li>User support contact</li><li>Documentation</li><li>References</li><li>Link to the resources from the OpenRiskNet library</li></ul> |

---

[2] https://openrisknet.org

An important part of the information related to the services is on the "training and user support" that integrates all available resources on the respective service (tutorials, video demonstrations, publications, etc.). To facilitate even more the use of the catalogue the categorisation shown below is used for filtering and selection of appropriate services by the users. The description of services is aligned to other indexing services like OpenRiskNet, the European Open Science Cloud (EOSC) and eInfraCentral templates, as well as to bio.tools, and automatic indexing in these services is anticipated in order to provide the largest visibility of the NanoCommons offerings possible. Figure 1 shows an example of the integration of the Jaqpot service into NanoCommons.
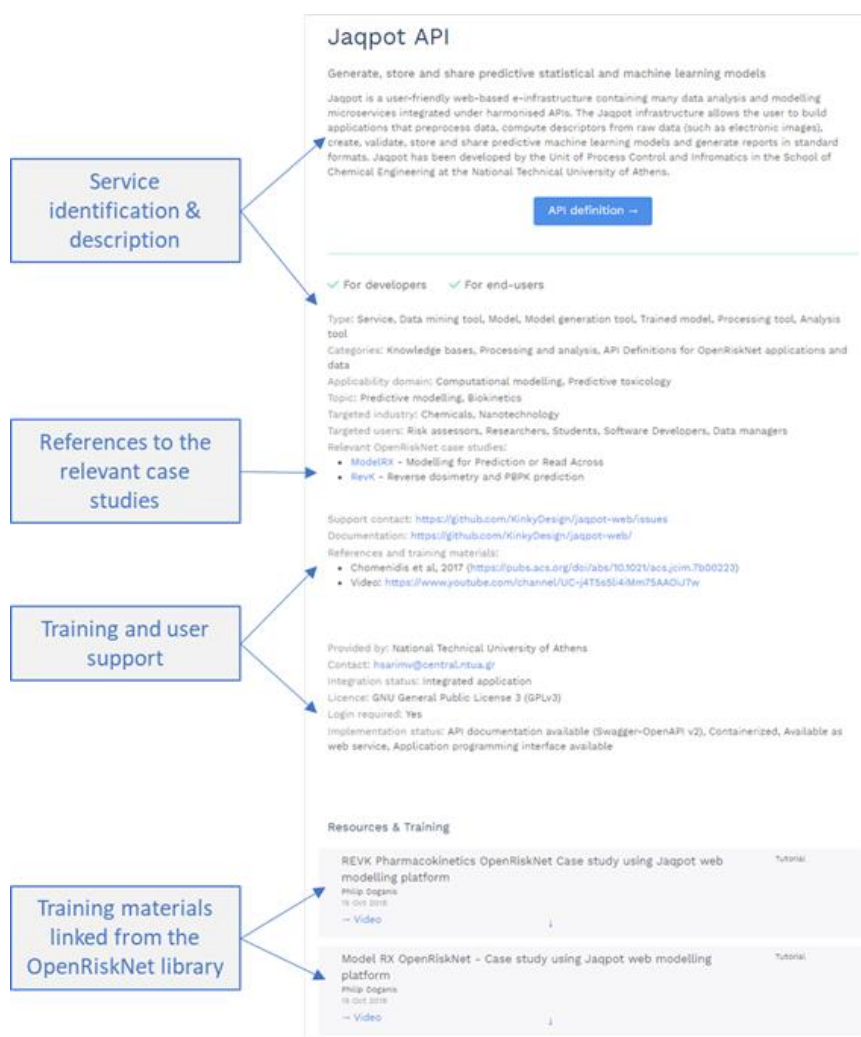


**Figure 1**. The Jaqpot service as an example for the integrated information and support system implemented in NanoCommons.

# 3. NanoCommons integrated services

## 3.1 Data management and sharing

NanoCommons is dedicated to the FAIR and Open data principles as embraced by the European Commission's guidelines for Horizon 2020 projects [2,3]. Based on these guidelines, the data offered through the NanoCommons KB and Data Warehouse (DW) need to be:
- Described using the necessary conventions and identified using rich metadata;
- stored in such a way that they can be accessed by humans and machines;
- structured in such a way that they can be combined with other data sets;
- licensed or have terms-of-use that spell out how they can be used by others.

Based on these principles, the NanoCommons e-infrastructure is providing nanosafety stakeholders (academia, industry, regulators) with the necessary tools and services to make the most impact with their research and innovation. NanoCommons acknowledges that FAIR and Open data are not simple checkboxes that needed to be crossed off, but rather are the basis for data innovation and socioeconomic impact [4]. Therefore, the principles are enforced on both existing databases like eNanoMapper, NIKC, NanoMILE, NanoFASE, and ACEnano, which are linked into the KB, and the NanoCommons DW which is responsible to publicly host the datasets curated and/or analysed within the NanoCommons KB that are not currently publicly hosted elsewhere.

In the initial version of the knowledge infrastructure the following DWs are integrated with ongoing efforts to improve their harmonisation and make them available via common search, browsing, and data enrichment functionalities via the semantic model:
- NanoCommons Data Warehouse
- NanoMILE
- NanoFASE Data Warehouses
- Nanomaterial-Biological Interactions Knowledgebase
- ACEnano Data Warehouse
- eNanoMapper and related project databases.

### 3.1.1 NanoCommons Data Warehouse

The NanoCommons DW is offering data management solutions to projects and individuals for storing datasets for which no more specialised solutions exist. It is fully integrated into the KB in the same way as the other databases for complete interoperability. This way, additional data can be made effectively open and reusable on demand, based on agreement reached between the NanoCommons project and the users as stated in the conditions of use of the NanoCommons DW.

To serve a multitude of different requirements, the NanoCommons DW implements a generic approach to data management. A number of semantic concepts such as "Nanomaterial", "Sample" or "Provider" are predefined in alignment with the requirements of the NanoCommons KB. However, data attributes for storage can be generated flexibly by the corresponding project (Figure 2). Description of those data attributes using the Nanoommons extension of the eNanoMapper ontology ensures that already existing attributes are re-used and, in addition, makes data findable and re-usable (the F and R of FAIR).
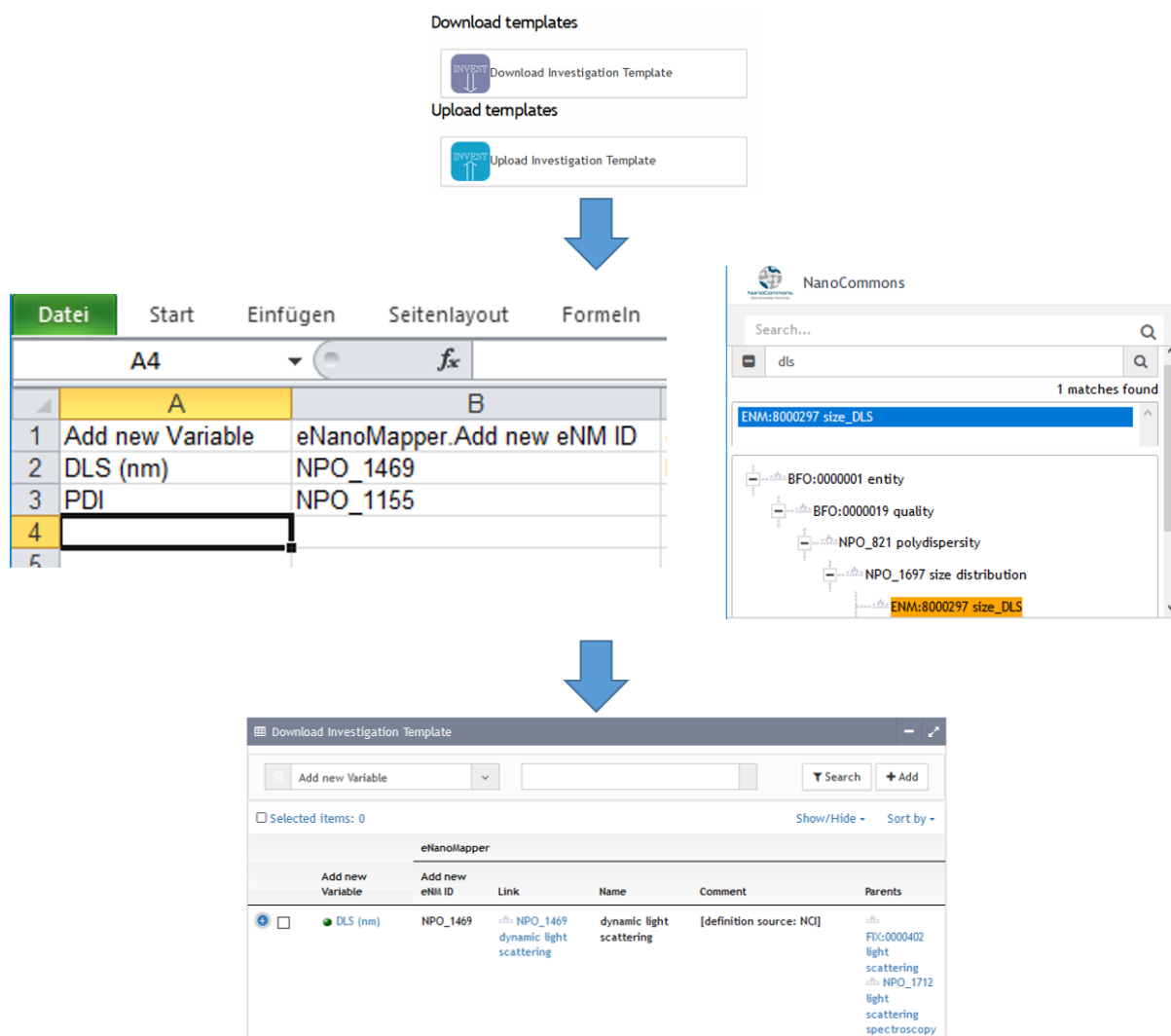
**Figure 2.** Templates (Excel) for data attribute generation can be downloaded from the NanoCommons DW. The "Investigation template" requests a description of data attributes in accordance with the eNanoMapper ontology terms (a corresponding eNanoMapper term search is provided on the NanoCommons DW).
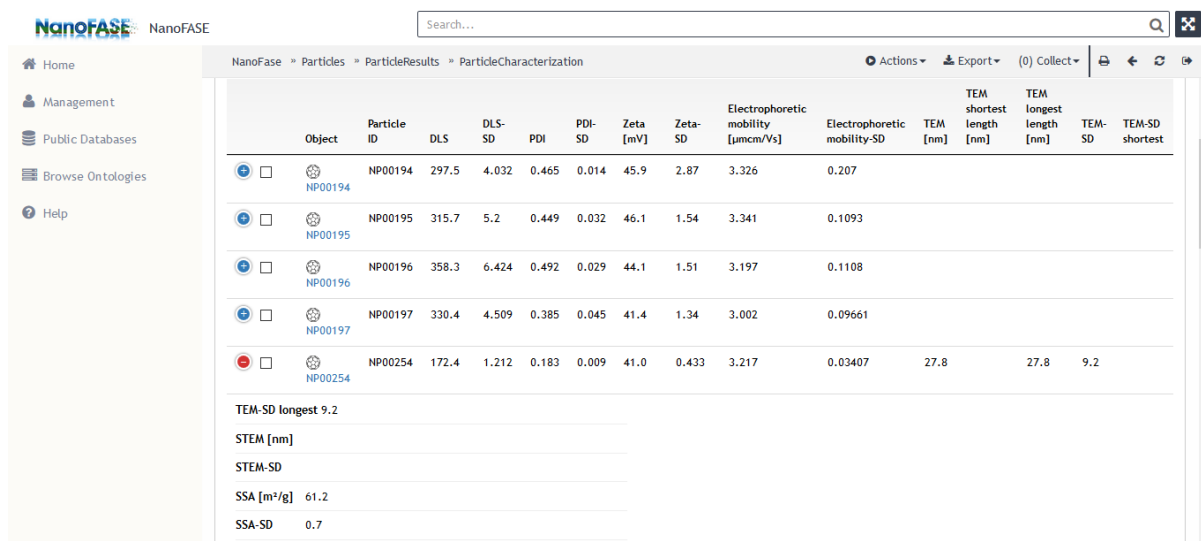
Project data is initially private until reviewed for data quality in order to avoid the distribution of incomplete/wrongly described data attributes or data. Integration of the NanoCommons DW into the NanoCommons KB makes the integrated data attributes and data available by API, providing standard format access and governance, following the principles laid out in the NanoCommons data management plan, such as accessibility and interoperability (completing the A and I of FAIR).

### 3.1.2 NanoMILE and NanoFASE Data Warehouses

NanoCommons will provide the nanosafety community with detailed physicochemical characterisation of the NMs used during the Framework Program 7 (FP7) NanoMILE and H2020 NanoFASE projects. These projects used (or aimed to use) a wide range of NMs (NanoMILE: 100, NanoFASE: 85) to test their effects on biological organisms and their environment fate and

transformations, respectively, and the underlying mechanisms driving these effects. The DWs are based on the same Biomax technology as used in the NanoCommons KB and contain information on a wide range of physico-chemical characteristics of the NMs (Figure 3) including:

- NM size
- Hydrodynamic diameter
- Polydispersity index
- Zeta potential
- Electrophoretic mobility
- Specific surface area
- Geometric surface area
- Energy band gap.



**Figure 3**. The Biomax NanoFASE KB containing the physicochemical characterisations of the project's NMs as determined for the NMs following production.  Many of the NMs were further characterised in the various test media representative of the different environmental compartments, and in a time-resolved manner, and these data are also captured, and linked to the pristine NMs dataset.
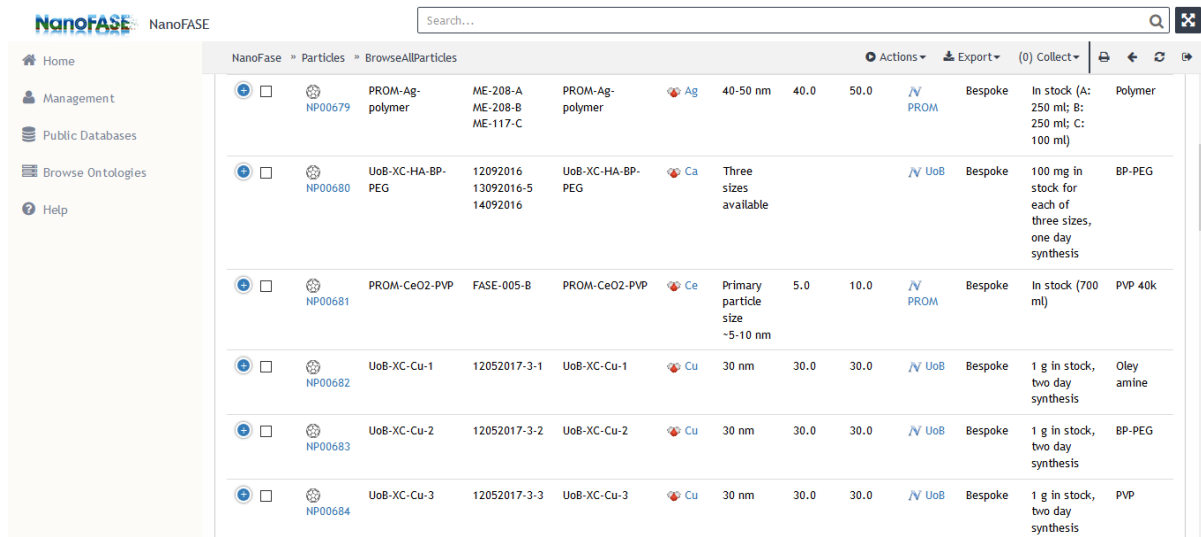
The provided metadata include (Figure 4):

- The supplier along with contact details and whether is a commercial or bespoke particle
- The chemical formula and Chemical Abstracts Service (CAS) numbers
- The identifiers used and can be included into publications/reports
- The nominal size (i.e. the size provided by the supplier)
- The NM phase (powder, suspension)
- The protocols and methods used for synthesis and physicochemical characterisation. These will be provided through the respective protocols repository either through the Biomax KB or the SciNote electronic notebook (ELN). For further information see section 3.2 on data management.

Besides the NM physicochemical characterisation, NanoMILE data on *in vitro* human cell lines (e.g.

A549 adenocarcinomic human alveolar basal epithelial cells, HCT116 colon cancer cell) dose-response toxicity tests, corona and uptake will also be available. NanoFASE data on NMs transformations in different environmental compartments and uptake into biota in these compartments is also included, along with the modelling data produced.



**Figure 4**. The NanoCommons KB containing the general metadata information of the NanoFASE NMs.

### 3.1.3 Nanomaterial-Biological Interactions Knowledgebase (NBI KB)

NanoCommons Users will also have access to the NBI KB built and maintained by our partners Oregon State University. The NBI is intended to offer industry, academia, the general public, and regulatory agencies a mechanism to rationally inquire for unbiased interpretation of NM exposure effects in biological systems. It serves as a repository for annotated data on NM characterisation (purity, size, shape, charge, composition, functionalisation, agglomeration state, Figure 5), synthesis methods, and NM-biological interactions (beneficial, benign or deleterious) defined at multiple levels of biological organisation (molecular, cellular, organismal). NBI also offers the possibility for visualisation of the hosted data (Figure 6).

The NBI provides structured annotated metadata (Figure 7) to facilitate data interoperability and maximise the impact of the data provided. The metadata include information on the NM used, the organisms, the media, exposure routes and the correlations made. Information and contact details are also provided on the data submitters and owners, allowing for direct communication between Users and experimentalists if further information/clarifications are needed.

**Figure 5**. The NBI knowledge base provides characterisation information on a broad range of NMs.

**Figure 6**. The NBI provides data and data visualisation on the dose-response and the subsequent biological effects (e.g. mortality, genetic effects) on several organisms (e.g. *Danio rerio*).

| Experiment #1 | | |
| --- | --- | --- |
| Field | Value | Download Datasheet |
| NBI Experiment ID | 200 | |
| NBI Material Identifier | 205 | |
| Exposure Metric / Assay | EZ Metric | |
| Primary Exposure Route | dermal | |
| Primary Exposure Delivery | Waterborne | |
| Secondary Exposure Route | oral | |
| Secondary Exposure Delivery | | |
| Tertiary Exposure Route | | |
| Tertiary Exposure Delivery | | |
| Study Factor(s) | concentration | |
| Exposure Organism | zebrafish | |
| Exposure Organism Life stage | embryo | |
| Exposure Length Classification | sub-chronic | |
| Duration of Exposure (hours) | 112 | |
| Investigation Title | MWCNTs OD: 20-30nm | |
| Investigation Description | Toxicological comparison between CNT diameter | |
| Investigator Name | Stacey Harper, Joseph Pryor | |
| Investigator Affiliation | Oregon State University | |
| Investigator Email | stacey.harper@oregonstate.edu | |
| Study Start Date | 2014-02-03 | |
| Exposure Organism Gender | mixed | |
| Exposure Organism Average Weight (mg) | 1 | |
| Exposure Organism Initial Age (hours post-fertilization at start of exposure) | 8 | |
| Continuity of Exposure | Continuous | |
| Exposure Temperature (Celsius) | 23 | |
| Exposure Media | Fish water | |
| Media Composition | 0.1% instant ocean salts | |
| Media pH | 7.1 | |
| Material Zeta Potential in Media (mV) | -15.4 | |
| Stable Average Agglomerate Size in Media (nm) | 96.53 | |
| Stable Agglomerate Size in Media Minimum (nm) | | |
| Stable Agglomerate Size in Media Maximum (nm) | | |
| Nanomaterial Preparation | diluted stock 3wt% solution with fishwater to desired exposure concentrations | |
| Experimental Notes | | |
| LC50 (ppm) | | |
| NOAEL (ppm) | | |
| Weighted EZ Metric EC50 (ppm) | | |
| Additive EZ Metric EC50 (ppm) | | |
| Publication Authors | | |
| Publication Year | | |
| Publication Journal | | |
| Publication Title | | |
| Publication DOI | | |

**Figure 7**. The NBI provides detailed annotated metadata on the performed experiments.

### 3.1.4 ACEnano Data Warehouse

Currently, this data warehouse supports the activities related to collection of physicochemical data and method optimisation in ACEnano project and aims to further disseminate this knowledge to the nanosafety community in a re-usable format. It includes multiple modules (protocols, data and workflows) to optimally accommodate the requirements of the different data types (e.g. raw and processed data, and protocols). The protocols repository facilitates adding, sharing and comparison of methods in a *questionnaire-like format* that guides the user through the documentation process from sample identification and description to the sample preparation, measurement and data processing (Figure 8). Storing protocols in a structured way rather than in an open text format allows for an easier comparison of the experimental design and how small changes to the protocol affect the final results which will be extremely useful for identifying method-related differences versus inherent variability in NMs batches. Furthermore, the DW offers long-term storage of the results in a re-usable format that are directly linked to the methods applied and the instrument metadata.
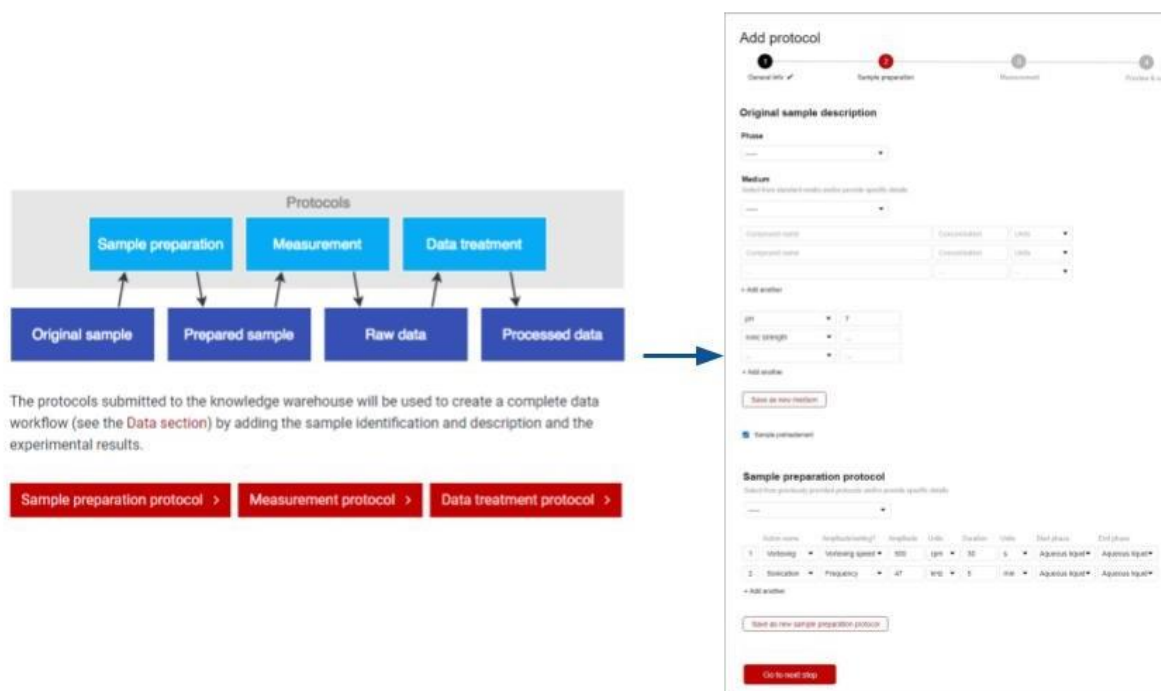
**Figure 8**. Structured protocols supported by ACEnano DW

### 3.1.5 eNanoMapper platform

In collaboration with OpenRiskNet, a Docker version of the eNanoMapper database software has been made available[3] [5]. The running instance of this OpenRiskNet service is an empty database, but data can be uploaded into this database in various formats. Guidance has been written outlining how data can be uploaded in a FAIR Linked Data format [6] and guidance from the eNanoMapper project has been updated and more widely disseminated, using the ELIXIR TeSS platform (https://tess.elixir-europe.org/content_providers/nanocommons).
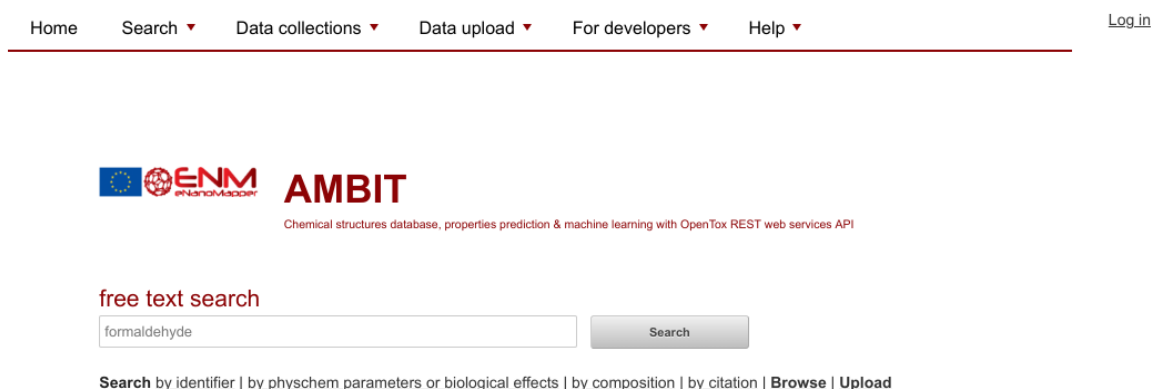


**Figure 9**. eNanoMapper DW running on the OpenRiskNet cloud.

---

[3] https://openrisknet.org/e-infrastructure/services/134/

The graphical user interface of the Docker is the default and depicted in Figure 9. The functionality is a FAIRer instance of the software released at the end of the eNanoMapper project. Since then, in collaboration with IdeaConsult Ltd (an eNanoMapper partner). The software has been improved with respect to the Linked Data input and output format [5]. This format is under development and extended following developments of the community nanosafety ontology ("eNanoMapper ontology") and described in this NanoCommons GitHub repository.

A second extension under development is the support of BioSchemas, in collaboration with the ELIXIR Toxicology Community and ELIXIR BioSchemas.org Interoperability Platform. BioSchemas is an extension of schema.org, a community standard initiated by Google and Microsoft for semantic annotation of websites, to support indexing by their Google and Bing search engines. The database software has seen the first patches to add support for the BioSchemas.org Dataset annotation, which is added when the HTML of a dataset ("bundle" in the eNanoMapper data warehouse terminology) is created. Because the name and description are later dynamically linked to the HTML, this needs to be added in a similar fashion to the BioSchemas.org annotation. This is currently lacking, causing the datasets to not show up in search engines yet. However, it is anticipated this will be implemented during 2019. The outcome will be a boost in the FAIR-ness of data hosted with the eNanoMapper platform. Furthermore, a BioSchemas.org extension for chemical substances is being developed, again in collaboration with thee ELIXIR BioSchemas.org Interoperability Platform, initiated at the first European BioHackathon organized by ELIXIR and held in Paris in autumn 2018.

One data set which is being FAIR-ified is around the toxicology of materials considered for nanomedicine. Katre Juganson *et al.* have published a paper in BJON about NanoE-Tox [7]. This database was composed by the Laboratory of Environmental Toxicology of NICPB (http://www.kbfi.ee/). The great news is that they released the content of this database under a CC-BY 2.0 license as supplementary information of the article. The bad news is that they released the data as an Excel spreadsheet, with a lot of the information being hard to process, and essential data being converted to dates (a common MS Excel problem [8]). The notebook to convert the data to RDF which can be read into an eNanoMapper instance can be found at https://github.com/NanoCommons/enmrdf/tree/master/NanoE-Tox. There is a manual step involved in the conversion, but most of the process is done by a script written in the Groovy programming language. Lines 19-29 convert dates back to NM sizes (see https://github.com/NanoCommons/enmrdf/blob/master/NanoE-Tox/convert2rdf.groovy#L19-L29), while the lines 31 to 189 basically describe ontology mappings for data in the spreadsheet (see https://github.com/NanoCommons/enmrdf/blob/master/NanoE-Tox/convert2rdf.groovy#L31-L189).

### 3.1.6 Other eNanoMapper-hosted data sets

Various other EU NanoSafety Cluster projects are using the eNanoMapper platform to make (mainly toxicological) data available in a FAIR way. This includes, besides eNanoMapper, data from the NANoREG, NanoReg2, caLIBRAte, GRACIOUS, and PATROLS projects, which is made available via various instances of the DW, indexed by this search engine: https://search.data.enanomapper.net/. However, only part of the eNanoMapper and the full NANoREG data is available under an open license. The other projects have not yet made their data available outside their own project consortia.

As soon as they make their data available under an open license (that allows resharing), taking advantage of the interoperability of the databases, data can be shared by other providers. At that moment, the data will be pulled into the NanoCommons platform.

## 3.2 Data management tools and workflows

The continuous advancement of technology for research purposes has impacted massively the way scientific research is conducted and the amount and quality of data produced. Experimental practices are becoming more and more complex, while the volume of produced data is increasing at an accelerating rate having surpassed the 9-year doubling cycle identified back in 2014 [9]. This is facilitated by the introduction of informatics and the shift towards translational research, big data analysis, machine learning and artificial intelligence analytical techniques which have introduced the need for the creation of larger, or multiple, datasets. For these techniques to be successful, the production and processing of high quality FAIR data is of the utmost importance.

To achieve this, streamlining everyday experimental practice, covering the whole data lifecycle (Figure 10) with specialised and detailed data management techniques that substantially decrease the time needed to gather, sort, manipulate and analyse experimental data is necessary. To this end, NanoCommons is integrating data management with experimental planning and data generation rather than data management being an add-on or post data generation tool. One means to achieve this is via the use of online laboratory notebooks (OLNs) in combination with premade and annotated data curation templates, which can facilitate research and is the first critical step to capture created data. The OLNs can also link, through local or wide area networks, from a few instruments to multiple laboratories around the world. In that sense, data harmonisation can be achieved along with the necessary metadata implementation, making the produced datasets suitable for assay/method development and further computational analysis and modelling.
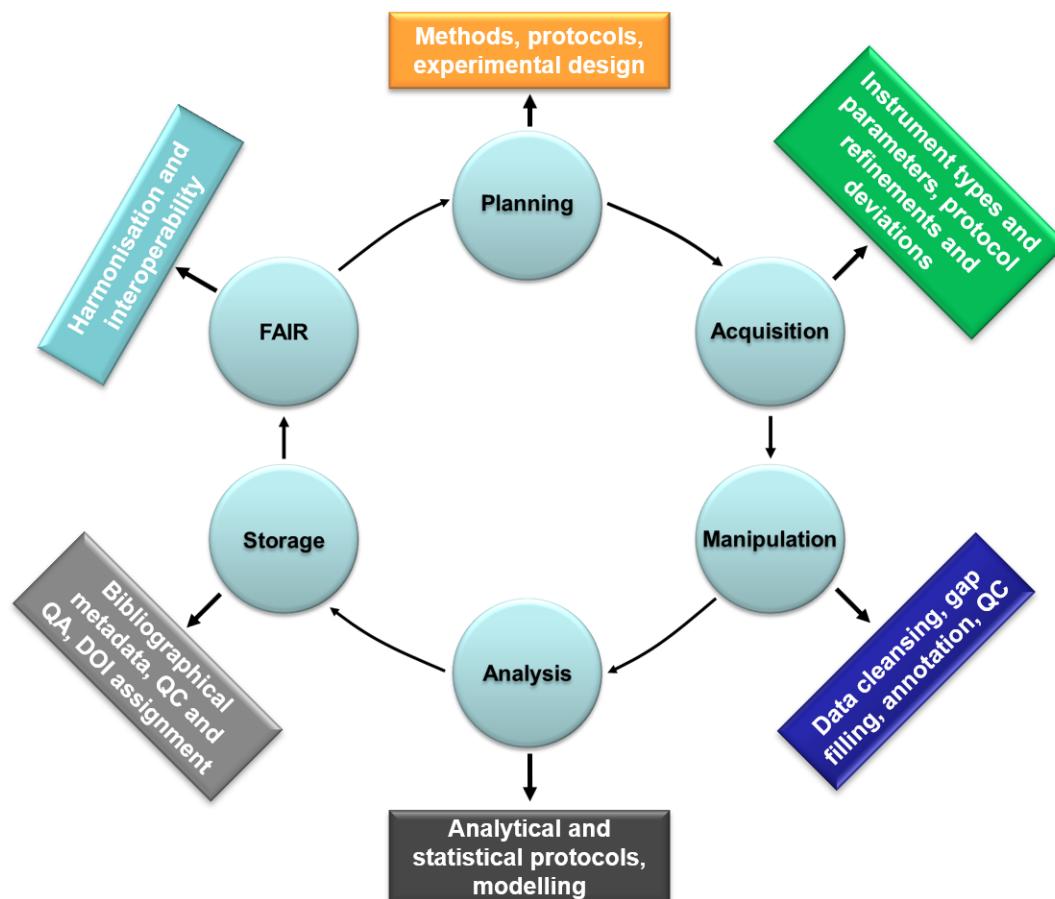
**Figure 10**. NanoCommons covers the entirety of the data lifecycle with specific tools that manage data and metadata acquisition, handling, analysis and exploitation.

### 3.2.1 Electronic Laboratory Notebooks (ELN)

NanoCommons offers a series of services that will assist with planning of experimental workflows (see section 3.3) and managing the produced and captured data. To begin with, NanoCommons offers the implementation of ELN aiming to facilitate experimental practice and data and metadata capturing as soon as it is produced. NanoCommons is using SciNote, which is an open access software allowing both local (https://github.com/biosistemika/scinote-web) and online (https://scinote.net/) installation based on the Users' specific needs, but integration of other OLN platforms can also be facilitate if user demand is there.

SciNote (Figure 11) provides the User with a flexible environment that allows the building and/or implementation of experimental workflows of increasing complexity. These can be confined locally into a single laboratory (Figure 12) or expanded to include laboratories and/or data analysts all around the world (Figure 13). The use of SciNote removes the need for data and metadata to be saved separately and shared within an experimental group, national or international consortium, thereby decreasing the possibility for data to be lost, misplaced or scattered and increasing the data and metadata quality and security.
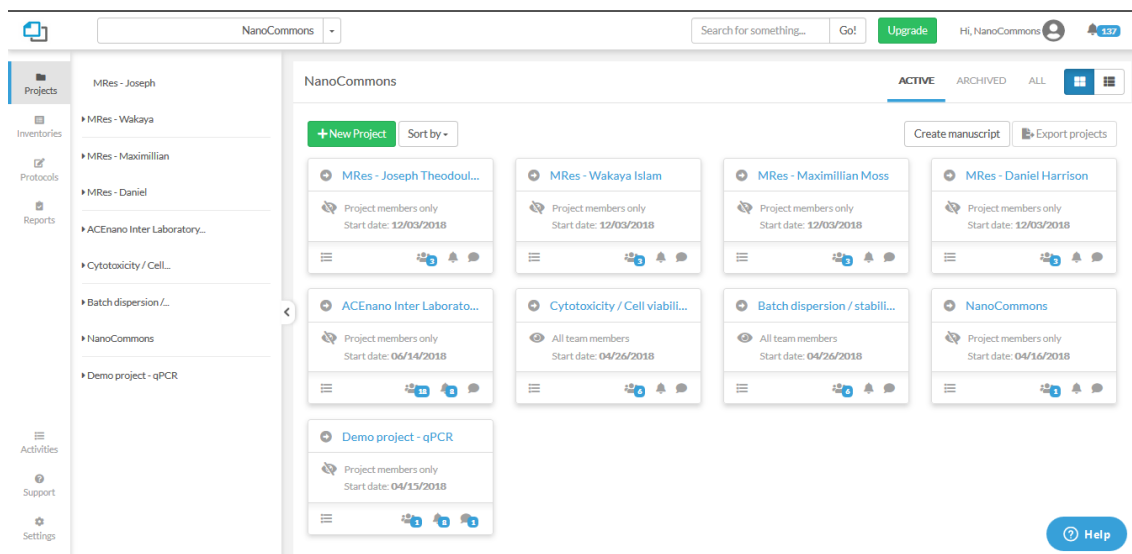
**Figure 11**. SciNote provides a flexible environment containing a dynamic experimental workflow area, along with a protocols' repository, custom data curation template implementation and automated reporting.



**Figure 12**. SciNote experimental workflow developed internally at the University of Birmingham for the physicochemical characterisation of NMs.
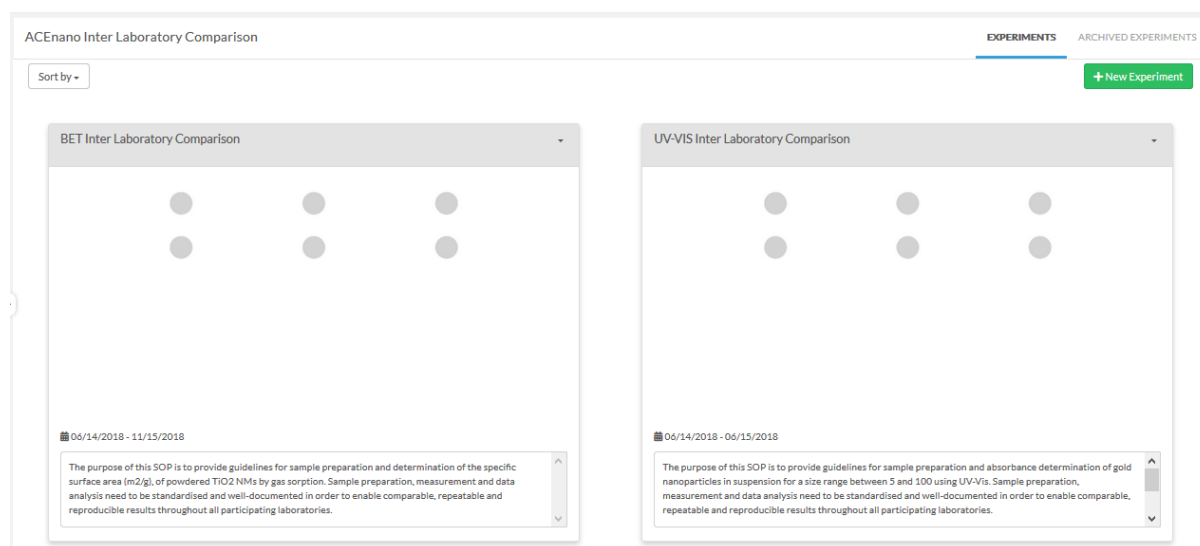
**(a)**
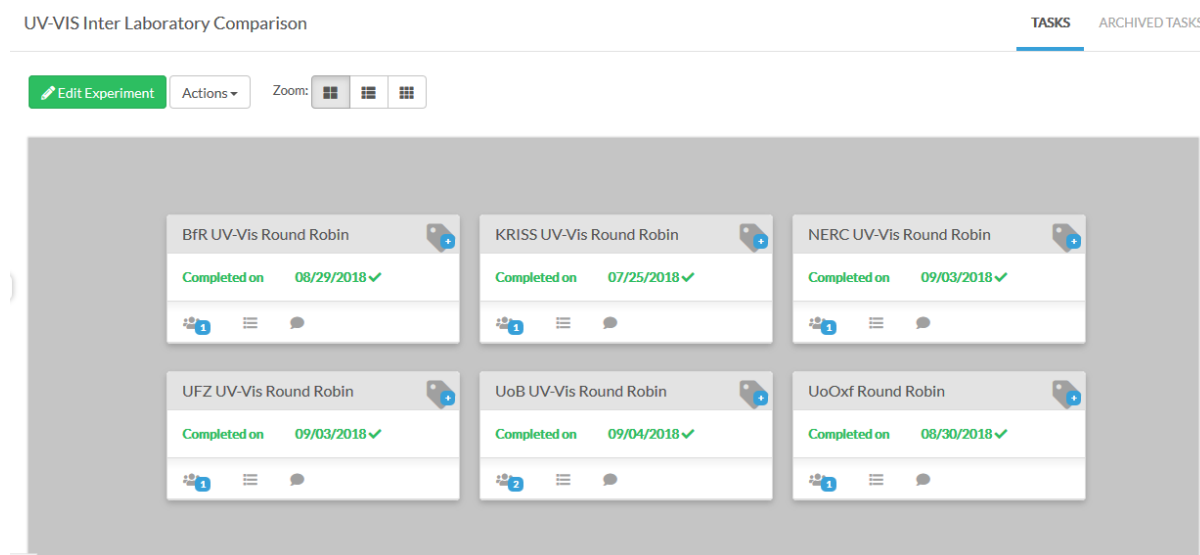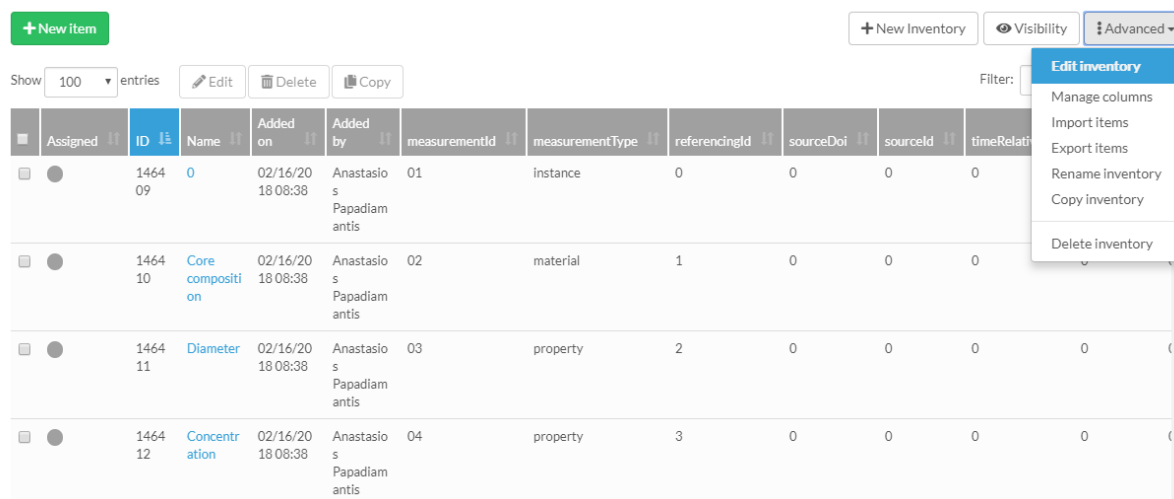


**(b)**

**Figure 13**. a) SciNote experimental workflows developed for the H2020 ACEnano project inter-laboratory comparison experiments, and b) specific partner assignments.

### 3.2.2 Data curation templates

SciNote offers the possibility to implement precompiled data curation templates to capture the produced and processed data and metadata (Figure 14). These templates are easily customisable to fit specific experimental needs and to be automatically extracted and mailed to the desired DW for uploading and storage. Within the NanoCommons TA approach, the data curation templates are prepared in collaboration with the NanoCommons experts following discussion with the Users to identify their specific requirements and experimental complexity. NanoCommons focuses on using two data templates:

A. the Nanoinformatics Knowledge Commons (NIKC) initially developed by our partners in the Center for the Environmental Implications of Nanomaterials (CEINT) of Duke University and modified by UoB and Biomax to meet the European experimental needs and guidelines; and

B. the NANoREG template [10], developed by the FP7 NANoREG project widely adopted within the nanosafety community.



**Figure 14.** The SciNote inventory (data curation template) with a NIKC template implemented.

NIKC is a dynamic and flexible data curation template, which meets the fundamental belief of the NanoCommons project that the NMs are inseparable from, and dynamically responsive to, their environment. This has often been described as NMs having both intrinsic (inherent to the material and unchanged irrespective of surrounding temperature, ionic strength, pH etc.) and extrinsic (context dependent) properties, where the latter are very sensitive to the surrounding conditions (pH, ionic strength and composition, available biomolecules, etc.). To this end, the need to characterise NMs under the exposure conditions, as well as before, during and after exposure has long been recognised in the scientific community, and is increasingly encoded in regulatory practice including the Organisation for Economic Cooperation and Development (OECD) Test Guidance and Guidance documents [11, 12]. To promote this best practice in the scientific community, complete monitoring of both the NM and the environment will be encouraged at every experimental stage, and is built into the preloaded templates. The NIKC can be used for both raw experimental and already published and curated data. This template introduces the concept of an *experimental instance* (Figure 15), which translates to every significant time point of the NMs' lifecycle, e.g. at synthesis, upon dispersion, upon contact with a test system (whereby the acquired biomolecule corona may exchange and evolve thus also altering the surround biofluid composition), at each sampling/measurement timepoint, following uptake into an organism, etc. The NIKC is an Excel-based complex template, which allows the simultaneous logging and reporting of both the experimental data and metadata (e.g. methods, instruments, images) information, and its EU NanoCommons modified version also ensures the automatic annotation with appropriate ontologies for direct insertion of the data/metadata into the NanoCommons KB.
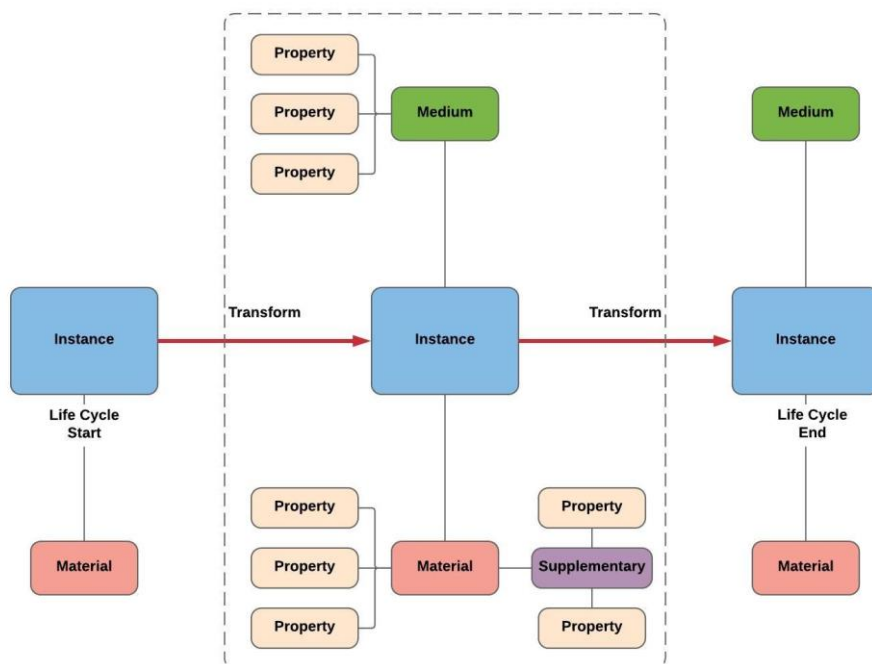
**Figure 15.** The experimental instance, which contains the monitoring and analysis of both the material and the environment in which it exists. Image courtesy of CEINT.

At the same time, the increasing complexity and size of experiments (e.g. mesocosm experiments whereby multiple end-points are studied across multiple species over extended time periods) present the need for dynamic and versatile templates that are able to accommodate different types of end-points and deviations from the original experiment design (Figure 16) or can link back to a common starting point (e.g. the pristine NM characterisation or the time 0 characterisation upon initial dispersion of the NM in the test medium). The NIKC offers that functionality with a tree-like structure, the trunk of which is comprised of the initial (pristine) NM(s) physicochemical characterisation. Following this concept, the NIKC data-tree can contain several branches that translate to various experimental variations (Figure 16). The NIKC approach also treats the experimental protocols and instruments both as data and metadata and links them automatically to the experimental data.

While the NIKC approach is primarily focussed on environmental transformations of NMs, it is equally applicable to NMs transformations during human and animal toxicological studies (e.g. blood corona formation, biodistribution, degradation in lysosomes, uptake of released metals into existing biochemical processes such as Fenton reactions etc.) and thus the experimental instance approach is being implemented for human/animal exposure also in NanoCommons, to enrich datasets already captured in flat databases such as eNanoMapper.

**Figure 16.** The NIKC is a dynamic and versatile data template that allows experimental deviation, while maintaining the underlying relationships and linking back to the original NM(s) characteristics. Image courtesy of CEINT.

The NANoREG templates are simpler premade templates based on the ISA-TAB-Nano format, which are free to use and cover a range of experimental procedures:

- Physicochemical characterisation
- Mammalian toxicology – *in vivo*
- Mammalian toxicology – *in vitro*

The NANoREG templates are easy to use, and according to the creators, can be modified to fit diverse experimental needs. One of the main advantages of the NANoREG templates is the harmonisation of the nomenclature and endpoints used with that of the OECD, which allows the curated data to be directly usable in a regulatory context. In this way, they offer an approach to handle meaningful and comparable data, a key success factor for the assessment of the quality and reliability of methods, as well as for an effective and responsible use of NMs. Their simpler form allows them to be used when the increased complexity of the NIKC is not needed, which also results in decreased cost in terms of working hours. The templates have been built by experts in different

fields (phys-chem, *in vivo* and *in vitro* toxicology) and are aimed at harmonising the logging of experimentally-produced data, and include meaningful and detailed information to support causal correlation analyses, modelling and Safe-by-Design (SbD).

The NANoREG template structure (Figure 17) is divided into 4 parts. The first part, called the modules, defines the field of study (physicochemical characterisation, *in vivo* and/or *in vitro* assays) for which the template will be used. The second defines the specific Excel workbooks as per the OECD list of endpoints and includes separate sheets according to the techniques/assays used. If a user needs to add a new technique/assay, they will need to add a new workbook sheet. Finally, each technique/assay is defined by the necessary experimental parameters (data and metadata). For the biological endpoints a section for the physicochemical characterisation of the material(s) used is included, which should be the characterisation performed in the specific exposure medium.
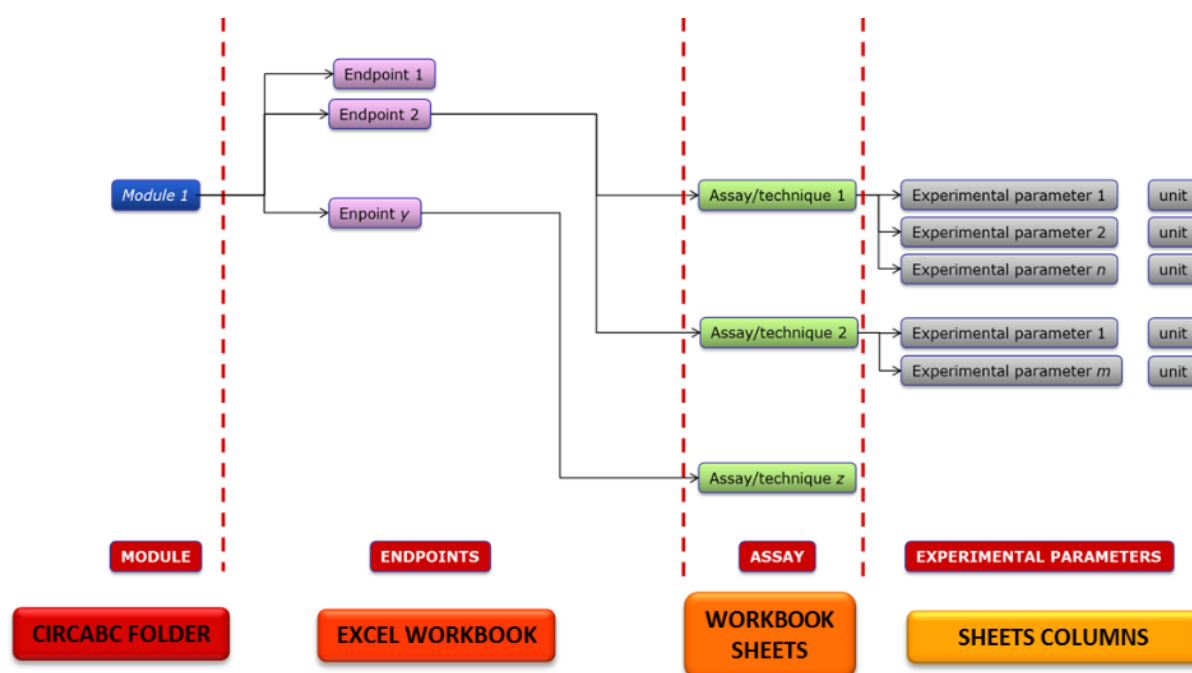


**Figure 17.** The NANoREG template structure.

### 3.2.3 Semantic annotation

To maximise the impact of the data included and to integrate diverse datasets into the NanoCommons KB a high degree of harmonisation and interoperability is required. For data interoperability to be successful, whether the data is Open or not, the data and metadata need to be semantically annotated using established ontologies containing clearly defined terms that fit the experimental, analytical, computational needs. NanoCommons is using established ontologies like the eNanoMapper ontology [13]. eNanoMapper is extensively used within the nanosafety community and is being maintained through our partners at Maastricht University (UM) that have a long history in making data interoperable.

The NanoCommons consortium acknowledges the significance of semantic annotation and is collectively working with Users and interested stakeholders to make their raw and/or processed data and metadata (academia, industry) and curated peer-reviewed published data FAIRer and when possible Open. The aim is to maximise the impact of data and promote data innovation, which is strongly interlinked with data FAIRness. To achieve this, the NanoCommons partners UM, DC and UoB will work closely with Users to identify the most appropriate ontologies and ontological terms to annotate their data and metadata. Another service will be to expand the eNanoMapper ontology with any missing terms either as standalone or through the combination of simpler terms to create more complex and more descriptive ones.

NanoCommons will use the tools developed by our partners Biomax Informatics for the semantic annotation of datasets. The Biomax annotation system (Figure 18) [14] provides a step-by-step guide allowing semantic annotation either through the uploading of a form containing the terms to be annotated along with the desired ontological IDs (Figure 19) or by using the Biomax Ontology Browser and annotation (Figure 20). Once uploaded along with the linked dataset, the Biomax system automatically annotates the dataset, makes it machine readable and harmonises it with other similar datasets increasing its interoperability and reusability.
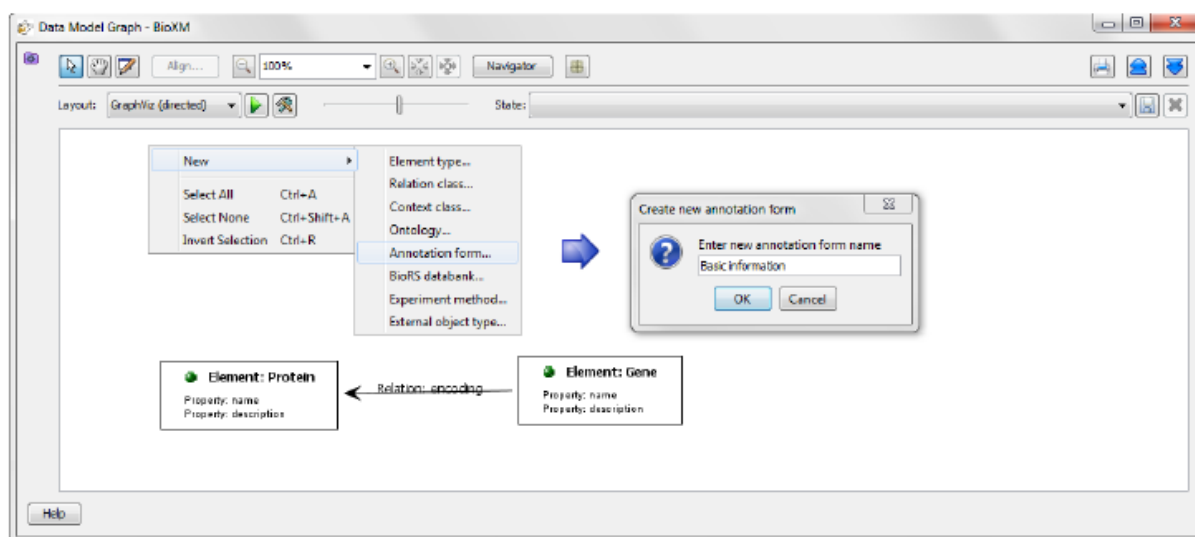


**Figure 18.** Term annotation during protocol and or data import.



**Figure 19.** Excel tab with term additions and ontological IDs. Once this is uploaded, the Biomax NanoCommons KW matches the ontological terms to the excel column terms in the protocol and to the resulting data, ensuring harmonisation.
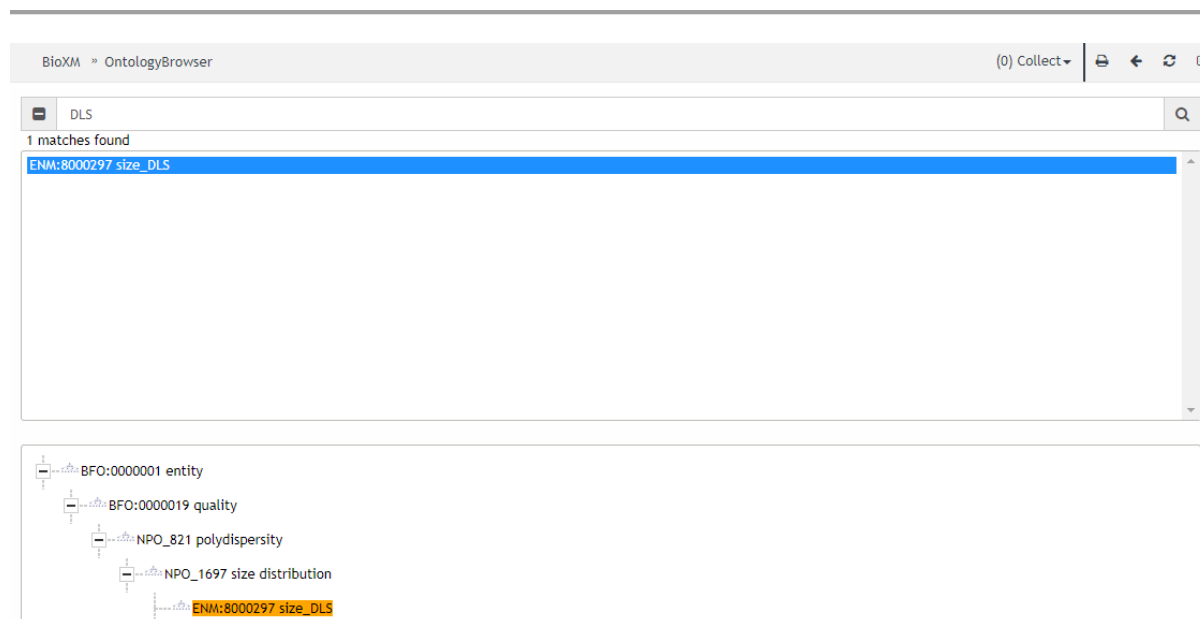
**Figure 20.** The BioXM ontology browser. User guidance for selection of terms and training via an ontology Hackathon is also available for users.

### 3.2.4 Protocols and methods repository

The data-heavy character of nanosafety research, its multi-disciplinarity spanning materials chemistry, biology, ecology, microbiology and pharmacology etc., and the rapid development of nanoinformatics have dramatically increased the need for high quality data [15]. The production of such data requires the close collaboration and feedback of all involved parties (data generators, curators, modellers etc.), and close alignment with established data reporting / minimum characterisation guidelines from each of the constituent fields [16–18]. The heaviest burden, though, falls upon the wet labs that need to produce a sufficient amount of high-quality, assimilatable NMs safety data by establishing streamlined experimental workflows to synthesise, characterise and assess the biological and ecological interactions of NMs, and to continuously demonstrate their reliability, repeatability and reproducibility. NanoCommons aims to support the generation of high quality nanosafety data through implementation of a protocol repository, protocol generation workflows and integration of protocols into the resulting datasets as part of the overall data packages generated and deposited into the NanoCommons KB.

Experimental protocols are the basis of data harmonisation and interoperability and need to be directly linked to the data they were used to produced. In this way, it will be possible to make the data FAIR, and to enable data querying and integration with other datasets originating from similar experimental workflows, and to analyse the effects of using different instruments or measurement principles to measure the same end-point and/or modifying experimental parameters within a protocol on the experimental outcomes.

Experimental protocols need to be well defined, provide all the necessary information and capture all necessary metadata, and where possible be standardised or validated through extensive testing. Experimental protocols need to provide experimental users with all the information needed to perform the experiment, and subsequent dataset users with all the information needed to interpret the resulting data. In the case of nanosafety protocols, a vital piece of information is the specific

media and/or NMs for which the protocol has been optimised, since NMs span such a wide range of chemistries and their physicochemical characteristics are highly dependent on their surroundings (so-called extrinsic properties). In most cases, though, it will be possible to modify existing protocols to cover a wide range of experimental parameters. In such cases, the protocols need to have built-in parameters such as medium and/or duration that can be modified and recorded accordingly, while keeping the method within the scope of the operational limits defined by the protocol validation. Potential deviations from the standardised experimental procedure need to be recorded and analysed to identify the effects of experimental deviations on the produced data (e.g. media of increasing complexity, extending the experimental duration etc.).

NanoCommons has established a specific workflow to be used for the implementation of all protocols into the NanoCommons protocols repository, which is hosted by our partners Biomax via the NanoCommons KB. This workflow is presented schematically in Figure 21 and includes the following steps:

   i.   Method identification
  ii.   General information (overview, health and safety, bibliographic metadata)
 iii.   Medium characterisation (where applicable)
  iv.   NM information (supplier, composition, concentration etc.)
   v.   Sample preparation protocol
  vi.   Instrument preparation (settings definition, calibration, equilibration)
 vii.   Sample introduction and equilibration
viii.   Settings refinement and measurement
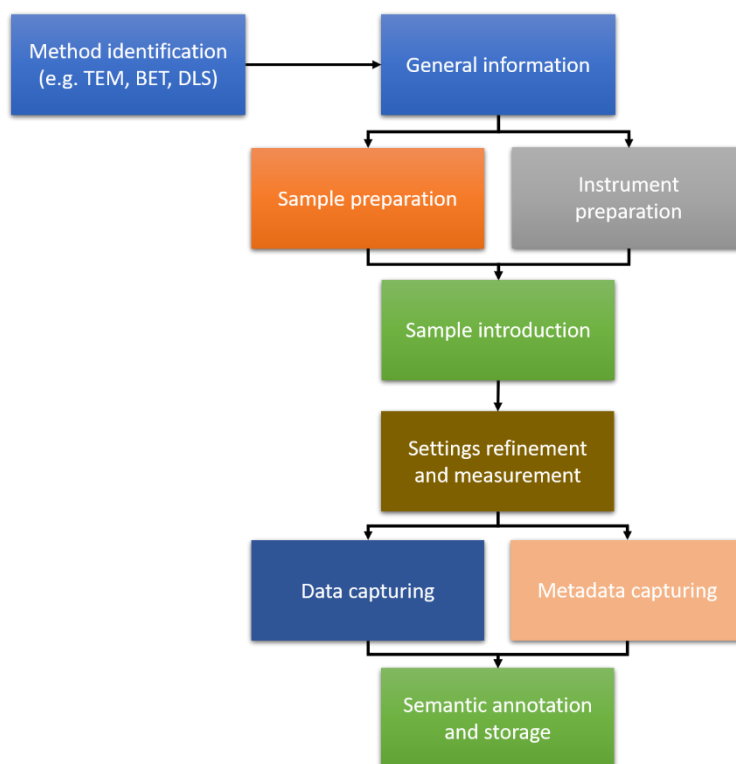  ix.   Data and metadata capturing and storage.



**Figure 21.** Protocol preparation workflow for submission to the NanoCommons DW.

The NanoCommons protocols repository will initially support all widely used formats of experimental protocols, i.e. text-based (Figure 22), Excel-based (e.g. NIKC instance protocols Figure 23), box-based workflow protocols (Figure 24) and Questionnaire-like format (through the ACEnano KW, Figure 25). Supporting these formats will allow the NanoCommons partners and Users to implement a continuously increasing number of protocols that were used or are being used in completed and ongoing projects, respectively, and link them to the respective experimental data. Eventually, NanoCommons aims to identify the optimum format that will facilitate protocol harmonisation and the ability to treat protocols as data points and convert all available protocols to these formats. In this way, it will be possible to analyse the effect of protocols' deviations to experimental data and promote standardisation and increase data quality.



**Figure 22.** Text-based protocol implemented into the SciNote protocols repository

| S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|
| parameterUncertaintyType | parameterUncertaintyLower | parameterUncertaintyUpper | parameterUncertaintyUnit | ownerGroup | protocolId | instrumentId |
| | | | | | [] | [] |
| | | | | | [] | [] |
| | | | | | [1] | [1] |
| | | | | | [] | [] |
| | | | | | [] | [] |
| standard deviation | -50 | -30 | millivolt | | [2,3] | [2,3,4] |
| | | | | | [] | [] |
| range | 6.8 | 7.2 | dimensionless | | [] | [] |
| | | | | | [] | [] |

**Figure 23.** The NIKC templates directly link the curated experimental data with the protocols/ instruments used. This is achieved using two reference columns protocolid (column X) and instrumentid (column Y) that link the data to the protocol and instrument utilised to generate it.
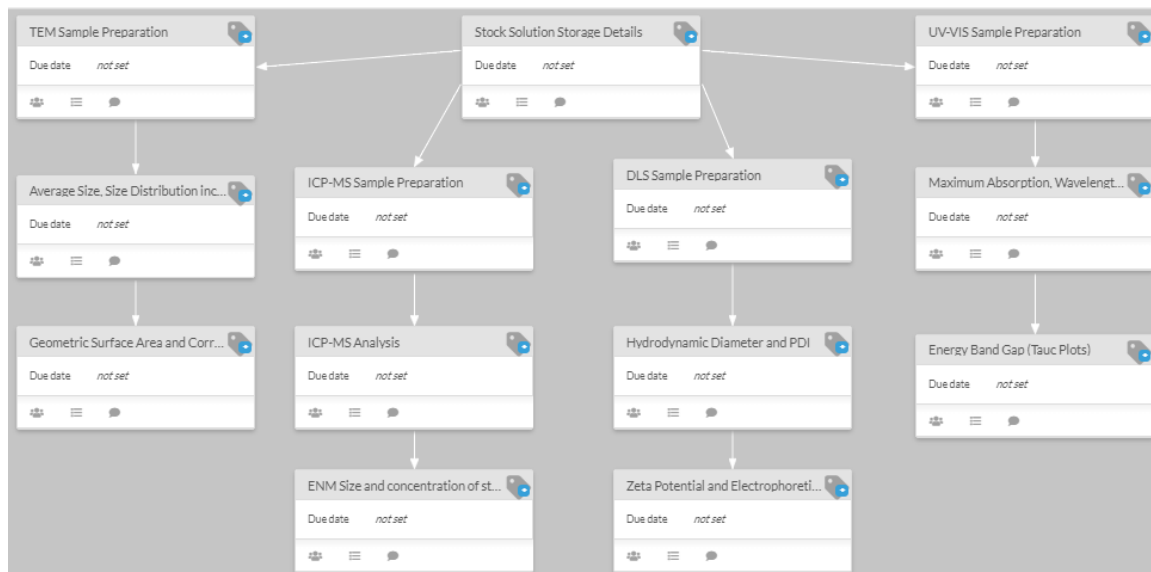
**Figure 24.** The box-based experimental workflow for the physicochemical characterisation of NMs as defined using the SciNote ELN.



**Figure 25.** An example protocol offered through the ACEnano KW.

The NanoCommons partners and users will be able to upload their protocols, following the process described above, directly into the NanoCommons KB using the Biomax uploading system, which also supports semantic annotation and enrichment, and the process depicted schematically in Figure 26. Protocols implementation can be achieved either through the upload of a file containing the protocol, or using the Biomax system to create custom experimental workflows (see Figure 27 for an example demonstrating the creation of box-based experimental workflows). The submitted protocols will be publicly available, unless they correspond to sensitive data and the development of novel experimental methods, with strict respect to the submitters' intellectual property rights. The protocols repository will also be integrated with the ELN used in NanoCommons (see section 3.2.1), to be able to be directly implemented and used with experimental workflows developed by the NanoCommons e-infrastructure Users.
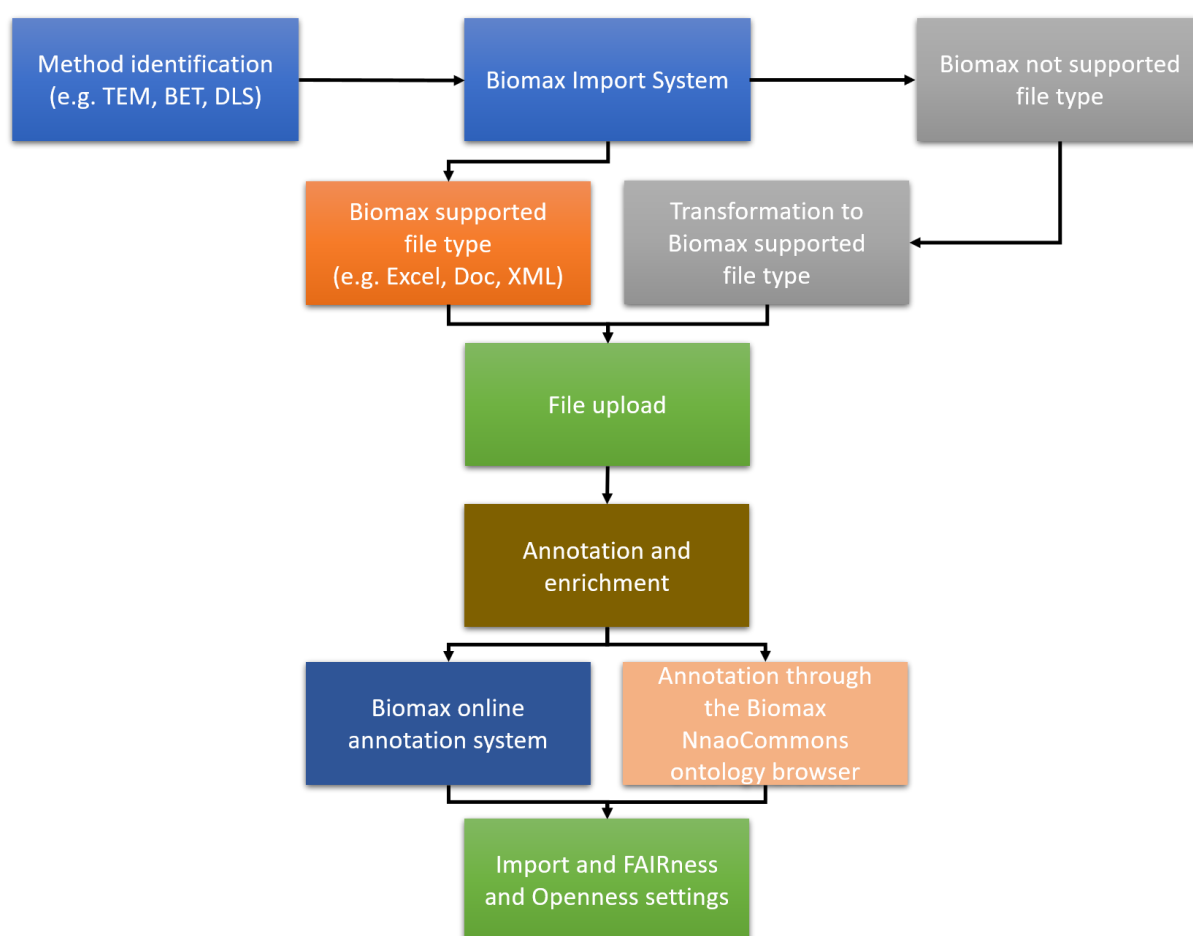


**Figure 26.** The NanoCommons / Biomax import system represented schematically.
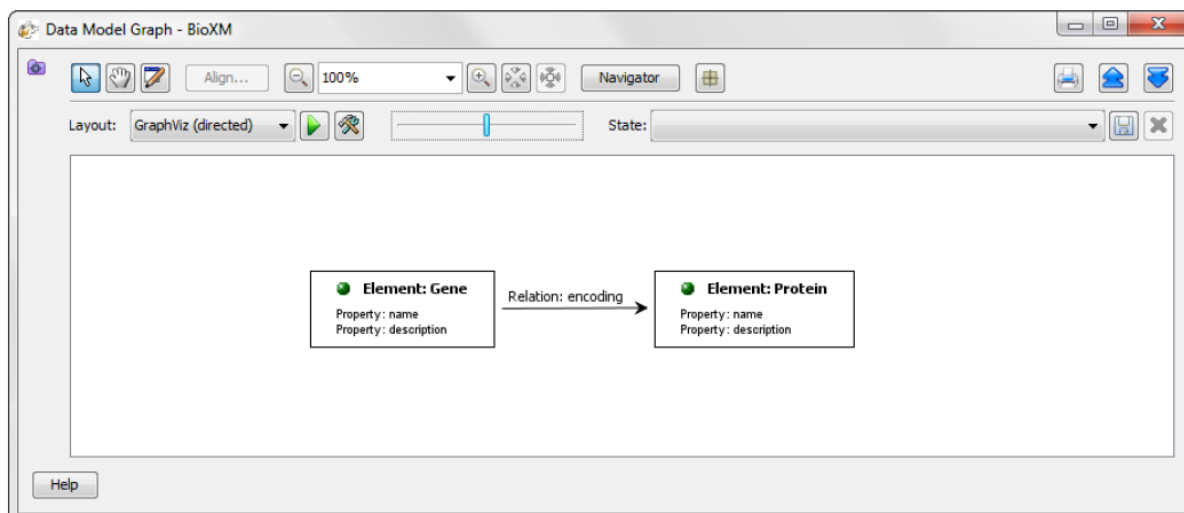
**Figure 27.** Example of a box-based creation of an experimental workflow and the relationship between different parts of it using the NanoCommons platform (developed by partner Biomax).

### 3.2.5 EdelweissData sharing platform

Similar to data storage, the NanoCommons data and protocol management concept is meant as the central entry point to obtain information on all available protocols and offers best practice data handling solutions for projects and individuals not having access to such procedures. However, individual projects may decide to establish such procedures and workflows first internally. To provide the Users with multiple options to select the most appropriate solution for their specific needs all harmonized and interoperable with the NanoCommons KB, the Edelweiss data sharing platform is offered as an alternative to the BioMX system. Edelweiss, developed by partner Douglas Connect, offers data management components that can be combined to form complete data solutions for internal or public sharing of data collections or for bringing single datasets into the NanoCommons ecosystem, preserving privacy and confidentiality if needed. The solution offers full interoperability with the semantic model of the NanoCommons KB and, in this way, allows for the full integration of the new data opening up all features of the KB like cross-database searching and browsing. In addition, the automatically generated APIs allow for linking of the data with the corresponding protocols as well as direct integration of the data into the analysis and modelling services of NanoCommons, or other external analysis and modelling tools. The ACEnano DW described above is the first example of a data sharing solution based on this technology and highlights the unique feature of intense linking of protocols, metadata and data via the questionnaire-like protocol and data upload workflows.

## 3.3 Development and implementation of experimental workflows

### 3.3.1 Experimental workflows

The need for high quality and interoperable data results in huge demand for detailed experimental workflows that ensure experimental completeness. The NanoCommons partners UoB, CEH and LEITAT support Users with experimental workflows that meet their needs. The NanoCommons

partners are liaising closely with interested stakeholders to identify their needs and help them design custom workflows to fulfil their scope, ensuring at the same time the high-quality nature of the data and metadata produced. To achieve this, NanoCommons organises site visits to either established facilities (as a reference in terms of layout, equipment needs, workflow, etc.) or to new facilities (in order to identify potential pitfalls). NanoCommons offers continuous remote support to Users during the design and implementation of the experimental workflows and throughout the entire data lifecycle, and documents the adopted approaches on the KB as best-practice examples and guidelines for other laboratories. Additionally, NanoCommons is creating web-based information exchange forums hosted on the KB, led by LEITAT, for NanoCommons experts and Users to exchange opinions and knowledge. Other services include the use of NanoCommons expertise to organise round-robin tests to evaluate inter-laboratory precision and standardise experimental methods and support on definition of testing workflows, particularly when parts of the process need to be externalised, consideration of what must be sequential versus what can be done in parallel, shipping, sample stability etc.

The expertise of the NanoCommons partners covers the full spectrum of nanosafety research:

- NM characterisation including transformations
- Nano-enabled product life cycle assessment
- Characterisation of nano-enabled products
- Quantification in complex matrices
- Preparation of NM suspensions for toxicological studies
- *In vitro* and *in vivo* NM toxicity studies including omics approaches
- NM ecotoxicity studies including environmental omics approaches
- Risk assessment of NMs.

### 3.3.2 NM physicochemical characterisation

The most important part of experimental workflows in nanosafety research is the detailed and high-quality physicochemical characterisation of NMs at various stages of the experimental process. UoB and LEITAT have a long successful record and experience in the synthesis and characterisation of NM that were used in several EU and national projects (e.g. NanoMILE, NanoFASE, NanoIsland). UoB holds a library of more than 100 unique instances of commercial and custom-made NMs.

UoB and LEITAT are using their expertise in experimental design in combination with the data management tools and services presented above to help Users design, implement and practice complex physicochemical characterisation workflows at various stages of the experimental process, starting from the pristine ENM and moving into characterisation in complex media. The NanoCommons experts are training and guiding experimentalists to not only capture the produced data with any gaps, but also the respective metadata to promote harmonisation and interoperability and assist with the development of read-across approaches for NM. NanoCommons is also providing a service to other projects supporting the development of their Data Management Plans.

### 3.3.3 Toxicology and Ecotoxicology

The Ecotoxicology & Chemical Risk Group at the Centre for Ecology & Hydrology (CEH), UoB and LEITAT Technological Centre are important nodes within the European Nanosafety infrastructure providing expertise and facilities in support of ecological effect and risk assessment studies with NMs including detection of toxicity and biodiversity effects in soils and surface waters. CEH offers support in the design of optimised workflows and the underpinning sample and data management needs, to maximise the suitability of the dataset for implementation into the NanoCommons community database and to ensure that data is suitable for use in subsequent modelling tools. Services can be applied to existing experimental set-up(s) or can support experimental facility design and subsequent access to data processing services. UoB offer support for aquatic community (e.g. biofilm) and organism (e.g. Daphnia) studies, including chronic and multi-generational exposure studies and mesocosm studies assessing food-chain effects and multi-stressor impacts (e.g. NMs and climate change impacts). UoB and LEITAT both offer support for *in vitro* toxicology also, including high throughout approaches.
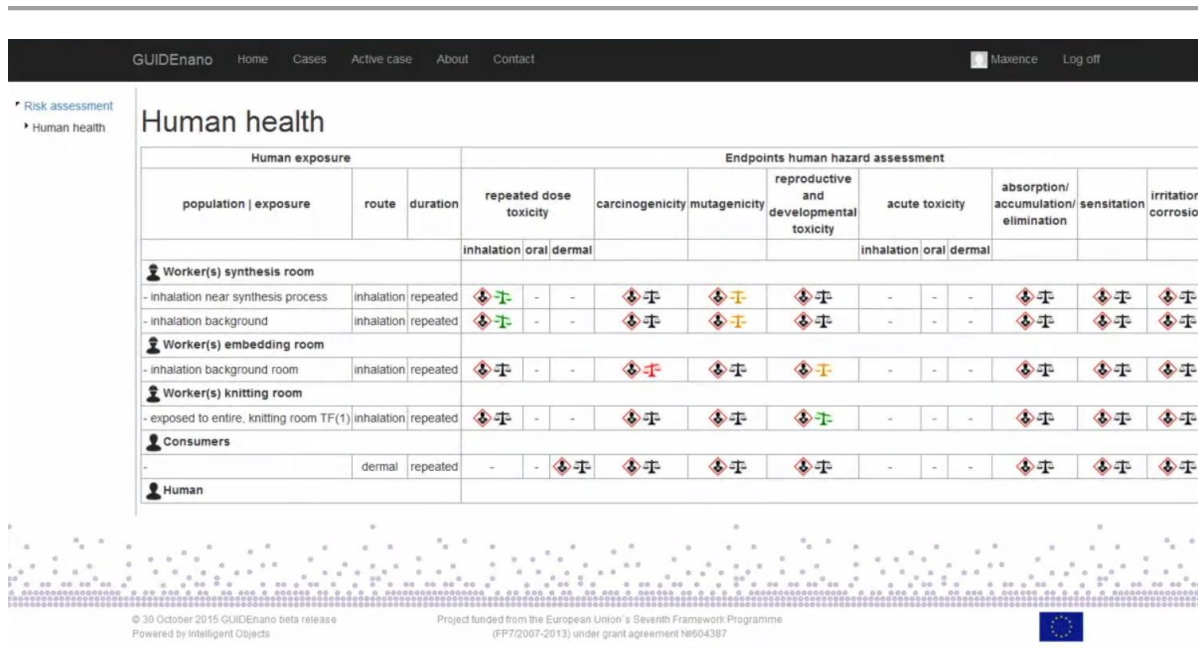
### 3.3.4 Biomolecule corona analysis (protein, lipid, carbohydrate, environmental)

UoB have extensive experience in terms of development of workflows for analysis of all aspects of the biomolecule corona, including the environmental corona composed of natural organic matter and secreted biomolecules including extracellular polysaccharides and proteins. We have recently developed protocols for analysis of the soil corona and the corona on NMs that have passed through the digestive tracts of organism and been secreted as pseudo-faeces and are subsequently available for uptake by other organisms. We have also been developing protocols for evaluation of the corona evolution over time following uptake into cells and organisms. The biomolecule corona is one of the context dependent parameters, and is largely responsible for recognition and uptake of NMs. The role of the small molecules, or metabolites, in determining the corona itself, and the consequences of the corona is an emerging area of research where UoB is leading, and is likely to provide important insights into NMs roles in AOPs.

## 3.4 Risk assessment and mitigation strategies:

### 3.4.1 Guidenano tool

The EU FP7 project GUIDEnano has developed a web-based guidance tool (Figure 28) for users to identify and apply the most appropriate risk assessment and mitigation strategy for NM-enabled products throughout their life cycle. The results obtained from the GuideNano project, as well as other past and ongoing projects and publications, have been incorporated into this tool, which guides the nano-enabled product developers (industry) into the design and application of the most appropriate risk assessment & mitigation strategy for a specific product. The full integration of the GuideNano tool into the NanoCommons KB, guaranteeing the correct implementation of this guidance, ensures that the risks associated with a nano-enabled product, throughout its whole life cycle, have been appropriately evaluated and mitigated to an acceptable level. The NanoCommons project, via our partners LEITAT, offers access and full training to the GUIDEnano tool along with continuous support for results optimisation. The GUIDEnano tool is also being integrated with the NanoCommons KB and KW, allowing decisions to be made based on ever more extensive datasets.

**Figure 28.** A snapshot of the GUIDEnano tool used for the human related risk assessment of NMs.

## 3.5 Omics data analysis

The integration of tools for big data (omics) analysis and mining into the NanoCommons KB serves two purposes. Firstly, it enables the identification of biological mechanisms and pathways associated with toxicity / adverse effects arising from exposure to NMs and produces aggregated biologically enriched descriptors. Secondly, many laboratories currently process and analyse their data in an insufficiently protocolled and irreproducible way. Standardising this processing, by providing integrated analysis tools with full capturing of metadata such as tool version, parameter settings, data version used etc. will help greatly with reproducibility of analysis and thus will increase the trust in the results.

The provided processing and analysis approaches are based on the integration of R tools, which have established themselves as the standard for the corresponding analysis approach in the omics data mining expert community. Standard workflows and default parameters based on published standards and dedicated research projects will be integrated directly into the user interface of the NanoCommons KB and, in this way, made available in a user-friendly, easy-to-use way. This will open the high-quality analysis to a larger target audience including non-expert users who will rely on pre-defined analysis methods and parameters without in-depth knowledge of data analysis approaches. TA Users can also access expert support for more advanced analyses and support with interpretation of the outputs from the models, as well as with optimisation of their datasets for use with these tools. The following areas will be covered in the KB. More information on the specific tools and their capacities are available in deliverable report D5.2.

- **Evaluation of data quality:** This function checks the selected data and gives recommendations concerning possible analyses based on the number of groups and replicates in the input data.
- **Normalisation:** To standard, well-established normalisation methods are offered: scale and Variance Stabilisation by Normalisation (VSN).
- **Differential expression analysis:** The R BioConductor packages limma, DEseq and edgeR are

used to assess differential expression by linear models. A linear model is fitted to the expression data of each probe/gene of a series of samples (with identical probes/genes).

- **Functional enrichment analysis:** The overrepresentation function uses the results of the differential expression analysis, applies a cut-off threshold for significance of differential expression and checks whether any functional parameter associated with the differentially expressed genes (DEGs), e.g. a Gene Ontology (GO) attribute or an Adverse Outcome Pathway (AOP), is significantly more often associated with a DEG than expected by chance. Fisher's exact test is used to assess overrepresentation in NanoCommons. To avoid predefined cutoffs, the Gene-Set-Enrichment Analysis (GSEA) searches for functional gene sets (e.g. a specific GO term, an AOP), which are enriched in the genes affected by a certain condition. All genes are ranked according to their degree of affection by the different conditions and all gene sets are then tested to determine whether the corresponding genes are enriched at one or the other end of the distribution. GSEA in NanoCommons uses the R BioConductor package limma and the rank function is romer (gene set enrichment analysis for linear models using rotation tests).

- **Network reconstruction:** Construction of Dependency Networks is achieved by using Mutual Information (MI), which is a measure of how well one variable (e.g. a gene) can be predicted by another (e.g. a second gene, a phenotype). In addition to standard correlation analysis, the concept of MI allows the identification of irreducible statistical dependencies that cannot be explained as an artefact of other statistical dependencies in the network. The R-package parmigene is used to infer the network based on the Kraskov Algorithm for estimation of MI and the ARACNE Algorithm for network reconstruction. Resulting networks can be queried and visualised directly in the NanoCommons KB or they can be exported for further analysis.

## 3.6 *In silico* modelling

The initial implementation of the Knowledge Infrastructure includes several computational services:

- **Jaqpot** (NTUA): a computational platform for *in silico* modelling of chemical compounds, which includes a variety of statistical tools for prediction of materials properties and risk assessment

- **NanoImage** (NTUA): set of tools for analysis of electron microscopy images of spherical particles and carbon nanotubes, allowing the user to derive descriptors for the NMs directly from images. Included in the Jaqpot package.

- **Enalos Cloud Platform** (NovaM): a cheminformatics and nanoinformatics platform with a variety of state-of-the-art modelling tools for the hazard prediction and risk assessment

- **NanoXtract** (NovaM): Nanoparticle image analysis tool for circular, cylindrical, plates and other NMs powered by Enalos cloud platform

- **Enalos Read Across Platform** (NovaM): zeta potential predictive model based on image nanodescriptors derived from NMs images

- **Materials Modelling tools** (UCD): set of electronic, atomistic and molecular simulation tools for in silico characterisation of NMs

- **SmartNanoTox bionano interface modelling tools** (UCD): evaluation of biologically relevant descriptors and activity of NMs.

### 3.6.1 Statistical modelling, data-driven NM property and functionality prediction, QSARs

NanoCommons offers TA to tools and services that allow the development of predictive nanoinformatics models and services based on statistical and machine learning methodologies and driven by available data. The produced models can then be shared with the community as ready-to use web applications. A brief description of the software used to create, validate and share predictive models is given below:

1. **Jaqpot:** Jaqpot is a computational platform for *in-silico* modelling of chemical compounds, which was originally developed by the National Technical University of Athens (NTUA) during the FP7 OpenTox project and further extended and continuously updated with new functionalities and services [19]. Emphasis is given on taking into account special requirements emerging from the complex and multi-perspective characterisation and evolution of NMs. Jaqpot allows the development, validation and sharing of nanoQSAR and biokinetics models, as well as read-across methodologies. All the functionalities are developed as microservices using the REST architecture. The various microservices are linked together through an extensive set of APIs which are documented with Swagger documentation and are available at https://api.jaqpot.org/jaqpot/swagger/ (Figure 29). The REST architecture and the APIs allow integration of Jaqpot with external services, including knowledge infrastructures and databases. For the end-users, Jaqpot services are offered through Graphical User Interfaces (GUIs). The present GUI can be accessed at http://www.jaqpot.org/ (Figure 30) and contains implementations of nanoQSAR models that have been presented in the literature. For example, Figure 31 presents a snapshot of the Jaqpot implementation of the linear nanoQSAR model for predicting log(1/LC50) to HaCaT for metal oxides [20]. A new Jaqpot graphical user interface is under development and can be accessed at: https://app.jaqpot.org/. It already contains nanoQSAR model implementation, including the example mentioned before (Figure 32). A variety of nanoQSAR models are now integrated in the Jaqpot platform and will be presented in detail in the NanoCommons deliverable report D5.4. The Jaqpot platform can be used by the community as a continuously updated repository of predictive models, ready to be used as user-friendly web applications.

**Figure 29:** Jaqpot API Swagger documentation
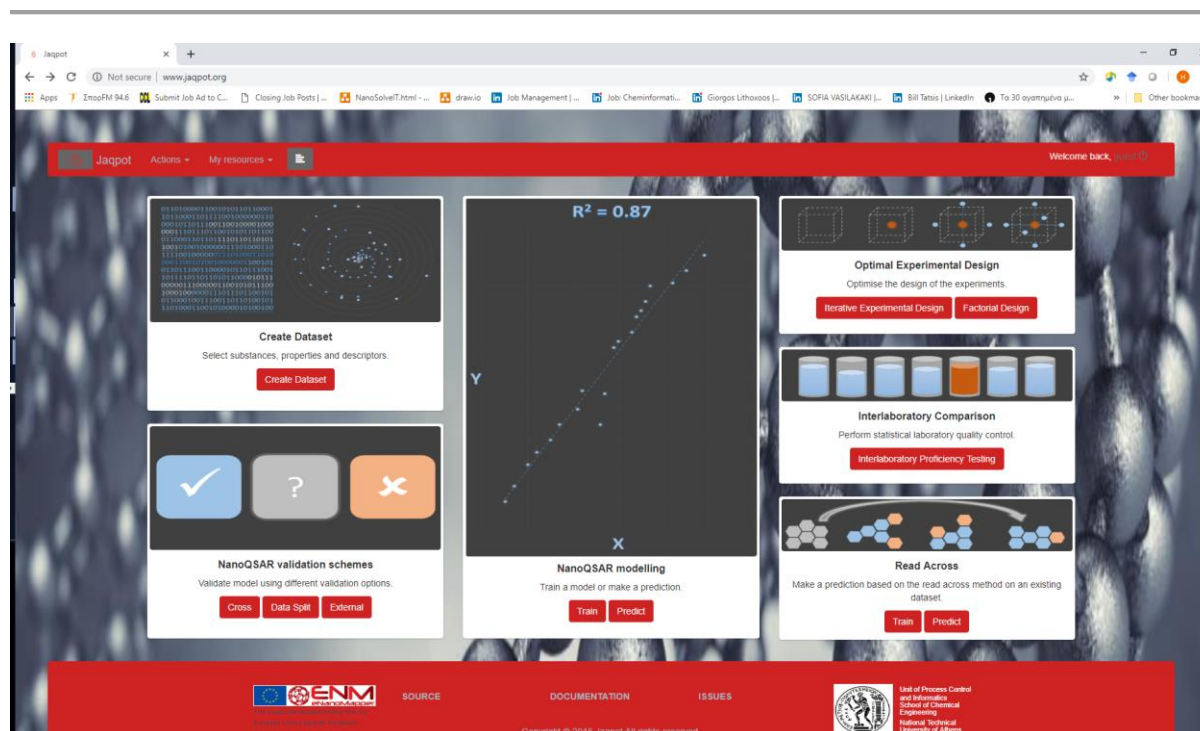
**Figure 30:** Home page of Jaqpot GUI
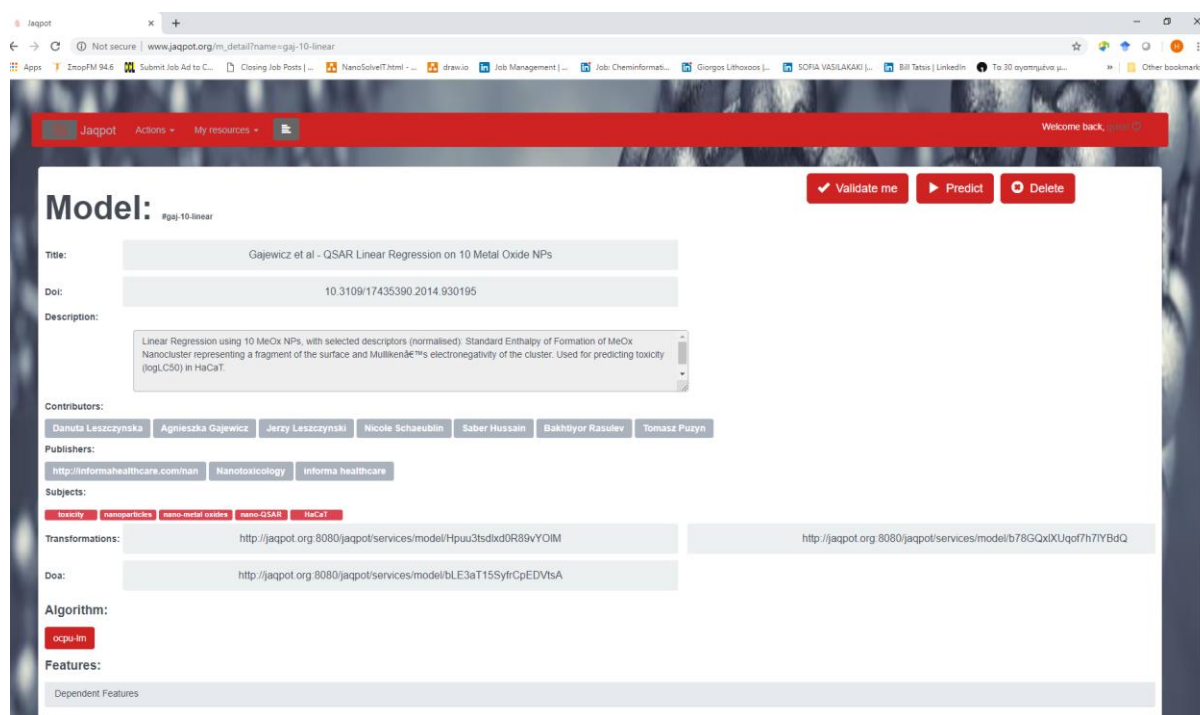


**Figure 31:** nanoQSAR model predicting log(1/LC50) to HaCaT for metal oxide NMs in Jaqpot.
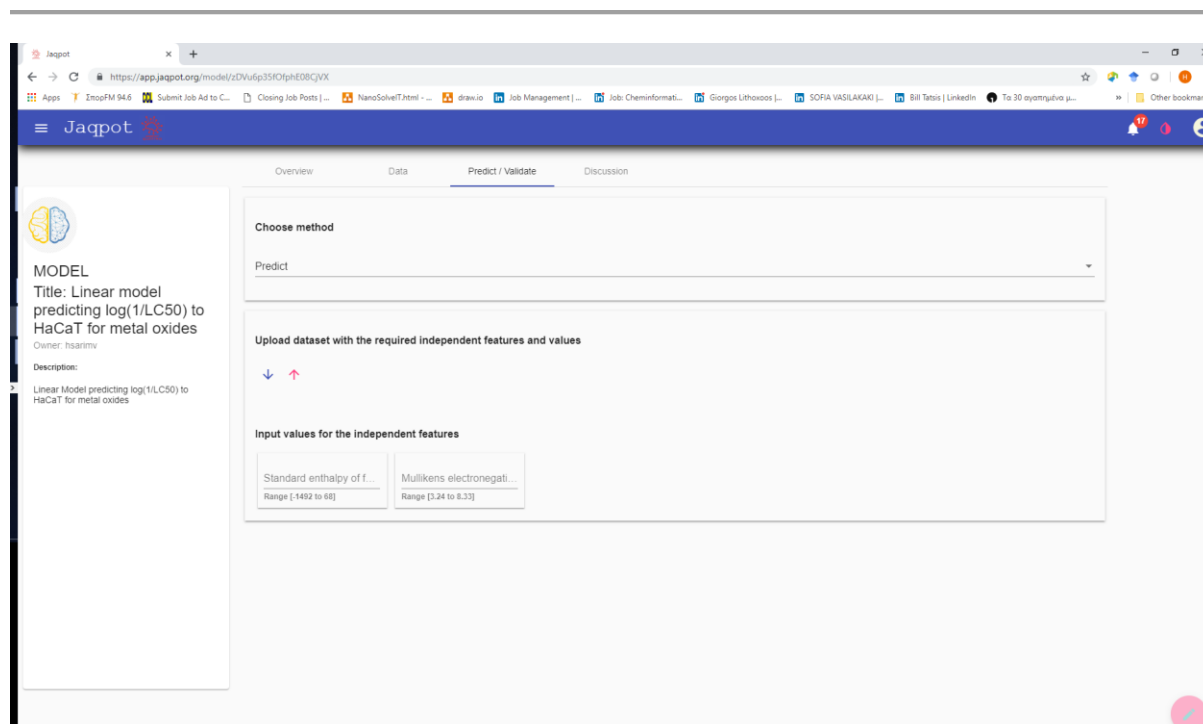
**Figure 32:** nanoQSAR model predicting log(1/LC50) to HaCaT for metal oxide NMs in the new Jaqpot Graphical user interface.

## 3.6.2 Materials modelling, physics-based NM property prediction

NanoCommons offers TA to a set of tools for *in silico* materials characterisation using physics and chemistry-based models, covering a wide range of length scales. These tools are developed to complement experimental characterisation of the NMs and quantify the properties of relevance for their biological activity. The electronic properties are evaluated with semi-empirical quantum calculations using the MOPAC package. Where necessary, the molecular and electronic properties obtained from semi-empirical calculations are complemented and verified by higher level of accuracy calculations, mainly, through density functional theory (DFT) methods and the NWChem package. For a few properties, e.g. hydration free energy, molecular adsorption, van der Waals energy, etc. molecular dynamics (MD) simulations with e.g. GROMACS are used. Furthermore, NanoCommons offers mesoscale simulation tools for evaluation of bionano interaction properties, such as protein binding affinity ranking with statistical models.

A brief description of the software used to evaluate the NM molecular descriptors is given below.

1. **MOPAC:** This semi-empirical program is widely used in the calculation of the geometries and energies of biological systems (proteins), crystals, co-crystals, polymers and other condensed phase systems. MOPAC 2016 [21, 22] has been improved with regards to the description of intermolecular interaction. Thus, the generation of electronic descriptors for QSARs is the canonical route. The effect of solvents can be captured by the COSMO method [23] where both ground and excited state geometries of solvated systems are optimized.

2. **NWChem:** This modern computational chemistry program package offers state-of-the-art methodology from quantum to classical and all combinations implemented for the efficient calculation of a variety of molecular and electronic properties of biomolecules, nanostructures

and solid state systems, utilizing parallel computing resources most efficiently [24]. Noteworthy are the multiple capabilities in the areas of kinetics and dynamics of chemical transformations, chemistry at interfaces and in the condensed phase.

3. **SIESTA:** This program package and method is used to perform efficient electronic structure calculations and *ab initio* MD simulations of molecules and solids [24, 25]. Localized basis sets and linear-scaling algorithms feature fast and accurate simulations. This approach projects the electron wavefunctions and density onto a real-space grid. In particular, layers of NMs of a specific size using periodic boundary conditions can be modelled and treated with density functional methods.

4. **GROMACS:** This versatile package performs MD simulation and energy minimization [26, 27]. The Newtonian equations of motion for systems with hundreds, to millions of particles are solved. GROMACS introduces hybrid acceleration by making use of graphical processing units to accelerate non-bonded force calculation.

5. **DFTB+:** This quantum mechanical simulation software package [28] is designed to run quantum mechanical simulations in an approximate way to DFT. The dynamics and electronic structure can be studied in molecules, solutions and solid-state materials.

6. Other software: Additional NM properties can be evaluated using free simulation packages OpenMM, LAMMPS[4] [29], GPAW. In **Table 2** all the descriptors are listed along with the methodology and software to be applied as well as the class of NMs that have/can be addressed and the nm size range which is reported in previous publications. The majority of the descriptors (band gap, ionization potential, density of states, electronegativity, etc.).

### 3.6.3 Interface properties

1. **ESPResSo** is a highly versatile software package for performing and analyzing MD many-particle simulations of coarse-grained atomistic or bead-spring models as they are used in soft matter research in physics, chemistry and molecular biology [30, 31]. It can be used to simulate systems such as polymers, liquid crystals, colloids, polyelectrolytes, and biological systems, such as proteins, DNA, and lipid membranes. It also has a dissipative particle dynamics and lattice Boltzmann solver for hydrodynamic interactions, and allows several particle couplings to the Lipid Bilayer fluid. ESPResSo is free software licensed under the *GNU Lesser General Public License*.

2. SmartNanoTox modelling tools (SNT-MT) are developed by H2020 project SmartNanoTox for prediction of the biological activity of NMs [32–34]. The tools include NM and biomolecule coarse-graining using Python scripts, evaluation of parameters of interactions such as potentials of mean force for biomolecular segments at the surface of NMs (using freeware Gromacs package), and prediction of protein 3D structure using free iTasser tool [35]. The considered biomolecular segments are as follows:

   ○ Amino acids side chain analogues. There are 20 naturally occurring amino acids from which all proteins of living organisms consist. Each full amino acid contains a peptide backbone fragment which is common for all amino acids. In order to avoid redundancy, we excluded the backbone fragment and considered the side chain analogues for all amino acids

---

[4] http://lammps.sandia.gov

excluding glycine (for which side chain analogue is just a H atom) and proline which has a different structure. This set of side chain analogues consists of 18 molecules.

○ Glycine and proline amino acids with the backbone fragment (GLY, PRO).
○ Modified charged forms of amino acids having pKa values between 4-10 (HIS+, GAN, CYS-)
○ Molecules representing fragments of most abundant lipid types: choline and phosphate group of phosphatydilcholine lipids; amino group of etanolamine lipids, ester group.
○ d-glucose as representing sugars.

The above list of molecules covers all the main types of chemical entities: hydrophobic, polar, aromatic and charged, and represents well all typical molecular fragments present in biofluids. A set of binding free energies of these molecules constitutes a simplest "fingerprint" of how an organism sees the NM and thus determines the biological response to a NM.

**Table 2**: Methodology and software tools for the calculation of theoretical descriptors of NMs

| Descriptor name | Method | Simulation package |
|---|---|---|
| Band gap | PM6, DFT | MOPAC, NWChem |
| Ionization potential | PM6, HF, DFT, TDDFT | MOPAC, NWChem |
| Density of states | PM6, HF, DFT | MOPAC, NWChem |
| Hydrophobicity | CG | SNT-MT |
| Dipole moment | PM6 (COSMO), DFT | MOPAC, NWChem |
| Atomic charge | PM6, DFT | MOPAC, NWChem |
| Total charge | PM6, DFT | MOPAC, NWChem |
| Electronegativity | PM6, DFT, LDM | MOPAC, NWChem |
| Absorption spectra | CIS, TDDFT | MOPAC, NWChem |
| Binding energies | PM6, DFT, DFTB | MOPAC, GPAW, DFTB+ |
| vdW energy | CG, PM7, DFT | SNT-MT, MOPAC, NWChem, SIESTA |
| Corona kinetics | CG | ESPResSo |
| Binding affinities | CG | SNT-MT |
| Hamaker constants | CG, Lifschitz theory | SNT-MT |
| Protein adsorption energies | CG, DFT, DFTB | SNT-MT, GROMACS, SIESTA, DFTB+ |
| Vibrational frequencies | PM6, MD | MOPAC, GROMACS |

## 3.7 Image analysis tools

### 3.7.1 NanoImage

The NanoImage tool, available at https://app.jaqpot.org/nanoImage/, is a tool for the analysis of Transmission Electron Microscopy (TEM) images, initially for spherical NMs, but within NanoCommons expanded to cover Carbon NanoTubes (CNT) as well. NanoImage produces 25 descriptors for spherical NMs and 6 descriptors for CNTs. Its functionality is available as both a Graphical User Interface and as a web service over an API, which allows its integration into the NanoCommons infrastructure. The Jaqpot Swagger documentation provides an easier way for developers to explore its capabilities. The source code is available at https://github.com/KinkyDesign/image-analysis.

### 3.7.2 NanoXtract

NanoXtract (http://enaloscloud.novamechanics.com/EnalosWebApps/NanoXtract/) is a unique online tool for the calculation of image descriptors based on TEM images of NMs developed for and made available through the Enalos Cloud Platform. The tool can generate a set of 18 image descriptors that greatly enhance and enrich the information extracted from images (currently size, size-distribution and qualitative information on shape and morphology), and which can be explored using predictive models (https://www.youtube.com/watch?v=Fl8qS-LdLUs) to identify those descriptors most predictive of NM behaviour and /or biological effects.

## 3.8 NanoCommons Helpdesk

NanoCommons aims to offer continuous support to its Users and to all stakeholders that would like to benefit from the services and tools the project offers. NanoCommons expert support will be offered through the NanoCommons Helpdesk, which has been established as part of the NanoCommons KB and can be accessed at www.nanocommons.eu/helpdesk. The main role of the NanoCommons Helpdesk is to ensure that the project will be targeting, integrating and connecting all end-user group requirements from academic and industrial research to risk assessment in industrial and regulatory settings, especially including SMEs. This way, it will be possible to provide all users with the support they need to access the integrated publicly available data and the state-of-the-art modelling and assessment capabilities in the NanoCommons KB. To facilitate users in starting to use the infrastructure and to provide ongoing support to all data, tools and services integrated and developed within the project, the Helpdesk provides supporting material and help functions including online tutorials for each of the different computational tools (described in more detail in the NanoCommons library section), as well as direct access to the experts through the Helpdesk to support Users in developing their NanoCommons Service or TA proposals.

Due to NanoCommons dealing with different types of stakeholders and/or services, the Helpdesk has a defined structure employing 5 levels of user assistance. These are based on the category of the user query and the specialised expertise needed to address a specific question. The different Helpdesk levels are:

1. **The NanoCommons Helpdesk Library (www.nanocommons.eu/library)**: It contains generalised

Frequently Asked Questions (FAQ) and guidance documents and tutorials on the services offered by the project. These documents are public and freely accessible through the NanoCommons website, and provide guidance and examples helping users to identify the service(s) required to address their specific needs. If the provided material is not able to address the needs of the users, they are redirected to the NanoCommons issue tracker through which they are able to communicate directly with the project.

2. **Communications team**: It is the second level of the NanoCommons Helpdesk and is based within our partner Biomax. The team is receiving any communication originating from the NanoCommons issue tracker and the established Helpdesk email (helpdesk@nanocommons.eu). The Communications team is then responsible to disseminate the requests to the most appropriate expert, based on the NanoCommons services description and guidelines put together from the project partners. If unsure, the team forwards the requests to assigned representatives of the University of Birmingham and Maastricht University, leading WP7, who decide the next steps.

3. **First Level Experts team**: The third level of the Helpdesk structure consists of experts from all the NanoCommons offered services. These are nominated from all partners and their remit covers a specific stakeholder (academia, industry, regulatory). Their main task is to provide guidance to potential users and help them determine the correct tools to address their need, and thus the best partner to work with initially and then the means through which the partners support user application development. They are also used as a first point of contact for the self-study users in order to refine the training materials and continuously develop the FAQ supporting materials for TA and Training.

4. **Second Level Experts Team**: This Helpdesk level consists of NanoCommons experts specialised in specific offered service(s). They are contacted if they are identified from the previous category as the most appropriate to provide help and guidance to users regarding their application or the use of a specific service/tool. They are responsible for any specialised training needed and to provide information on the functionality, limitations and the domain of applicability of the model and whether the prediction produced from the tool is robust. This is important for preventing non-appropriate use, which would lead to erroneous conclusions.

5. **Senior Experts Panel**: In extreme cases, when the identified experts are not able to answer a specific query, the question is redirected to the most senior scientists responsible for the specific research field. It is expected that these cases are extremely limited in number. In any case, NanoCommons needs to plan ahead for such cases. The responsibility of this group is to address any relevant issues, or to identify external experts that are able to assist with answering such scientific questions.

For the NanoCommons Helpdesk structure to be functional, the Communications Team needs to successfully:

1. receive, filter and distribute requests to the relevant specialist;
2. ensure that no request is lost and that all are dealt within a timely manner;
3. be able to evaluate requests and follow-ups for completeness of answers and response times in order to improve the workflow where needed.

To achieve this, a dedicated issue tracker system, Request Tracker (RT) has been set up, which is run

by NanoCommons partner Biomax. Users getting access to the issue tracking system following first contact with the Helpdesk to make sure it is not misused by spam-mail, i.e. users have to be registered to the issue tracker before being able to call it via email.

# 4. Conclusions

NanoCommons offers tools and services supporting the nanomaterial risk assessment and SbD communities in making existing data FAIR, implementing data management, mining, handling, analysis, visualisation and modelling workflows into their experimental design and data collection processes and to facilitate direct integration into relevant databases (experimental and *in silico*) with all of the require metadata and semantic annotation.

To facilitate and harmonize access, improve user experience and easiness to use as well as offer centralised support and access to expert knowledge, it has established the NanoCommons KB as the central one-stop shop to the NanoCommons services and is continuously extending the integrated offerings. All services currently available are briefly presented in this deliverable report, whereas more detailed information can be obtained in the individual deliverable reports of WPs 3-6 wherein the tools were developed or optimised for integration into the NanoCommons e-infrastructure platform. Additionally, this deliverable report described how Users can browse the KB to find services relevant to their work via the service catalogue and the Helpdesk features and then use these services as self-learners or apply for TA access providing expert-support and customization.

The first call for TA was launched on 4th February 2019, with 6 monthly calls for access planned throughout the period 2019-2021. Further tools and services will continue to be integrated into the catalogue as they become available, both from NanoCommons partners and from external providers who recognise the benefits of integration to promote utilisation of their tools and services by the nanoinformatics and nanosafey communities.

# References

1. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3: 160018.

2. Guidelines on FAIR Data Management in Horizon 2020 [Internet]. 2016 Jul. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3. Open Research Data and Data Management Plans [Internet]. ERC Scientific Council; 2018 Apr. Available: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

4. Oxford-Research, *Preliminary analysis: introduction of FAIR data in Denmark*. 2018, Danish Agency for Science and Higher Education. https://ufm.dk/en/publications/2018/preliminary-analysis-introduction-of-fair-data-in-denmark.

5. Danyel J, Bayjanov J, Farcal L, Oki N, Geerke D, Doganis P, et al. Report of the Service Integration with OpenRiskNet (Intermediate Report) (Deliverable 4.2). 2019; doi:10.5281/zenodo.2596999

6. Willighagen E. Adding nanomaterial data [Internet]. Available: https://tess.elixir-europe.org/materials/adding-nanomaterial-data

7. Juganson K, Ivask A, Blinova I, Mortimer M, Kahru A. NanoE-Tox: New and in-depth database concerning ecotoxicity of nanomaterials. Beilstein J Nanotechnol. 2015;6: 1788–1804.

8. Zeeberg BR, Riss J, Kane DW, Bussey KJ, Uchio E, Marston Linehan W, et al. Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics. BMC Bioinformatics. BioMed Central; 2004;5: 80.

9. Van Noorden R. Global scientific output doubles every nine years : News blog [Internet]. Available: http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

10. NANoREG templates for data logging – Towards a harmonised way of logging data within the NanoSafety Cluster community? [Internet]. [cited 18 Mar 2019]. Available: http://www.citeulike.org/group/19781/article/14026825

11. Deydier L. Updated REACH Guidance for nanomaterials: what you need to know (ECHA Webinar) [Internet]. 2017 Nov 30. Available: https://echa.europa.eu/documents/10162/23489227/301117_nano_webinar_environmental_deydier_en.pdf/7e3370b0-c40c-a8a9-bd40-5542bf9408c0

12. Stallmann M. First OECD Test Guideline on the investigation of nanomaterials [Internet]. Available: http://www.umweltbundesamt.de/en/topics/first-oecd-test-guideline-on-the-investigation-of

13. Hastings J, Jeliazkova N, Owen G, Tsiliki G, Munteanu CR, Steinbeck C, et al. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. J Biomed Semantics. 2015;6: 10.

14. Biomax, *BioXM™ Knowledge Management Environment — Quick Start*. 2017. http://www.biomax.com/lib/products/bioxm/Biomax_BioXM_Quick_Start.pdf.

15. Haase, Klaessig. EU US Roadmap Nanoinformatics 2030. 2018; doi:10.5281/zenodo.1486012

16. Marchese Robinson RL, Lynch I, Peijnenburg W, Rumble J, Klaessig F, Marquardt C, et al. How should the completeness and quality of curated nanomaterial data be evaluated? Nanoscale. 2016;8: 9919–9943.

17. Faria M, Björnmalm M, Thurecht KJ, Kent SJ, Parton RG, Kavallaris M, et al. Minimum information reporting in bio-nano experimental literature. Nat Nanotechnol. 2018;13: 777–785.

18. Stefaniak AB, Hackley VA, Roebben G, Ehara K, Hankin S, Postek MT, et al. Nanoscale reference materials for environmental, health and safety measurements: needs, gaps and opportunities. Nanotoxicology. 2013;7: 1325–1337.

19. Chomenidis C, Drakakis G, Tsiliki G, Anagnostopoulou E, Valsamis A, Doganis P, et al. Jaqpot Quattro: A Novel Computational Web Platform for Modeling and Analysis in Nanoinformatics. J Chem Inf Model. 2017;57: 2161–2172.

20. Gajewicz A, Schaeublin N, Rasulev B, Hussain S, Leszczynska D, Puzyn T, et al. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: hints from nano-QSAR studies. Nanotoxicology. 2015;9: 313–325.

21. Stewart JJP. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J Mol Model. 2007;13: 1173–1213.

22. Stewart JJP. Stewart Computational Chemistry, Colorado Springs, CO, USA, 2016. In: MOPAC2016 [Internet]. Available: http://openmopac.net/

23. Klamt A, Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2. 1993; 799–805.

24. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, Van Dam HJJ, et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. Comput Phys Commun. 2010;181: 1477–1489.

25. Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, et al. The SIESTA method for ab initio order- N materials simulation. J Phys Condens Matter. 2002;14: 2745–2779.

26. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008;4: 435–447.

27. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2: 19–25.

28. Aradi B, Hourahine B, Frauenheim T. DFTB+, a sparse matrix-based implementation of the DFTB method. J Phys Chem A. 2007;111: 5678–5684.

29. Plimpton S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J Comput Phys. 1995;117: 1–19.

30. Limbach HJ, Arnold A, Mann BA, Holm C. ESPResSo—an extensible simulation package for research on soft matter systems. Comput Phys Commun. 2006;174: 704–727.

31. Arnold A, Lenz O, Kesselheim S, Weeber R, Fahrenberger F, Roehm D, et al. ESPResSo 3.1: Molecular Dynamics Software for Coarse-Grained Models. In: Griebel M, Schweitzer MA, editors. Meshfree Methods for Partial Differential Equations VI. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. pp. 1–23.

32. Brandt EG, Lyubartsev AP. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO 2 (100) Surface. J Phys Chem C. 2015;119: 18126–18139.

33. Lopez H, Lobaskin V. Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. J Chem Phys. 2015;143: 243138.

34. Lopez H, Brandt EG, Mirzoev A, Zhurkin D, Lyubartsev A, Lobaskin V. Multiscale Modelling of Bionano Interface. Adv Exp Med Biol. 2017;947: 173–206.

35. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9: 40.