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 4.4

Deliverable	D4.4 First version of data warehouse & collaborative knowledge infrastructure
Work Package	WP4: JRA2 - Knowledge infrastructure
Delivery date	30 June 2019 (M18); Submitted 31 August 2019
Lead Beneficiary	Edelweiss Connect GmbH (EwC)
Nature of Deliverable	Demonstrator
Dissemination Level	Public (PU)

Submitted by	Dieter Maier and Ivan Stambolic (BIOMAX), Lucian Farcal and Thomas Exner (EwC), Anastasios Papadiamantis and Iseult Lynch (UoB), Egon Willighagen (UM)
Revised by	Antreas Afantatis (NovaM) and Vladimir Laboskin (UCD)
Approved by	Iseult Lynch (UoB)

This project has received funding from European Union Horizon 2020 Programme (H2020) under grant agreement nº 731032

Table of contents

Abbreviations

AOP - Adverse Outcome Pathway

API - Application Programming Interface

ECHA - European Chemicals Agency

ELN - Electronic Laboratory Notebooks

ERM - European Registry of Materials

FAIR - Findable, Accessible, Interoperable and Reusable

FAQ - Frequently Asked Questions

FP7 - Framework Programme 7 (2007-2013)

GDPR - General Data Protection Regulations

H2020 - Horizon 2020 (Framework Programme 2014-2020)

KB - Knowledge Base

KI - Knowledge Infrastructure

KW - Knowledge Warehouse

NC - NanoCommons

NC-KB - NanoCommons Knowledge Base

NM - Nanomaterial

PID - Persistent Identifier

PDI - polydispersity index

QC/QA - Quality Control / Quality Assurance

QSAR - Quantitative Structure Activity Relationships

RA - Risk Assessment

STEM - Scanning transmission electron microscopy

TA - Transnational Access

TEM - Transmission electron microscopy

Summary

Deliverable 4.4 is part of Task 4.2 that deals with the development of the NanoCommons Knowledge Base and Warehouse that aims to collect raw and processed data generated by different projects and to provide support and processes for preparing datasets for upload to the NanoCommons data warehouse or other specialized databases linked into the infrastructure, as well as templates for data collection (linked to the online notebooks). Additionally, repositories for protocol description directly linked to the relevant datasets are provided in order for complete coverage of the experimental procedure and results to be included into the system. The work in this task is based on the concepts and aligned with developments from previous and ongoing projects (e.g. eNanoMapper, NANoREG, NanoMILE, NanoFASE, SmartNanoTox, ACEnano, etc.) and is extended in order to cover additional areas of nano safety research. It considers requirements for regulatory reporting and Adverse Outcome Pathway (AOP) development, as well as the support of ontology development and semantic annotation. In this way, the warehouse facilitates data transfer to and from other databases as part of a federated data ecosystem.

The NanoCommons knowledge base (KB) specification and design was described in Deliverable D4.1, while here in Deliverable D4.4 we report on the first version of the KB implemented using the BioXM Knowledge Management Environment, the ongoing work to link it with other major data sources, access options to data via application programming interfaces callable by modelling and risk assessment tools as well as the NanoCommons service catalogue building the one-stop shop for finding integrated data and software resources.

Several aspects related to the KB user interface functionality are presented including the login and home pages, and example of functions for browsing the content of the warehouse. The work in this deliverable is directly linked to the ontology development (see Deliverable D4.3) and implementation, as the ontology integration allows users to quickly and easily search for ontological terms and identifiers that match the terms needed.

The description of the data integration from different data warehouses will focus on the first databases to be integrated and presents them and the associated challenges case by case: NanoCommons, NanoMILE and NanoFASE, eNanoMapper and its implementations for other projects (e.g. NANoREG) and ACEnano. Finally, the current implementation of the NanoCommons Catalogue of Services is exemplified.

Introduction

A key component of NanoCommons is its **data management system,** called the **NanoCommons knowledge base (KB)**, which is being built around appropriate ontological concepts developed in partnership with other major current projects (e.g., OpenRiskNet, NanoFASE, ACEnano) and building on the predecessor project eNanoMapper.

The KB specification and design was described in Deliverable report D4.1. Here, we report on the first version of the NanoCommons KB implemented using the BioXM Knowledge Management Environment, contributed by partner Biomax, and the ongoing work to integrate and interlink it with other major data sources to generate a global, universal repository for all nano safety related data and knowledge. Additionally, the proposed options for automatic access of the data by the NanoCommons modelling and risk assessment tools will be shortly outlined for consistency (more details can be found in Deliverable reports D4.2, D5.1 and D5.2) and the NanoCommons service catalogue is described being the one-stop shop to find integrated resources (data and software) and to get more information on those.

The NanoCommons KB consists of the specific semantic mapping of concepts and ontologies commonly applied in Nanoinformatics and Nanotoxicology research, the integration of data from important previous or current Nanosafety research projects such as NanoMILE, NanoFASE, eNanoMapper and ACEnano, the configuration of Queries and Reports for application programming interface (API) based integration of analysis and modelling tools and the configuration of a Browser based graphical user interface to access the semantically mapped information and facilitate the utilisation of the data visualisation and modelling tools. In this report, we describe:

1) How the user (as defined in D4.1, D10.3) can interact with the NanoCommons KB;
2) How the data input is supported by ontology search tools to facilitate the semantic integration;
3) How different existing databases are being or will be integrated;
4) How data access and download in formats suitable for modelling and input into the various visualisation and processing tools is streamlined;
5) How tools for specific exposure / hazard prediction and risk assessment (RA) tasks are managed so that they can easily be found by the users.

Knowledge Base user interface

In order to give an overview about how the NanoCommons Knowledge Base (NC-KB) user interface works, we provide in this section a short, screenshot based guide to the information available and the navigation within the user interface. More details on the the BioXM^{TM} software can be found in Losko and Heumann (2009) [1], while the method for generating the **semantic mappings** is described in Maier, et al. (2011) [2].

Login

Users access the login page to enter the secure, authentication and authorisation based NC-KB (**Figure 1**). This approach allows the NC-KB to be fully compliant with the General Data Protection Regulations (GDPR), and to provide tailored solutions and support for individual users. Currently the usage is restricted to NanoCommons partners. As a next step, an auto-registration will enable external users to automatically register and access the information and tools as soon as they are approved by NanoCommon partners for public access. Later stages may also implement single sign on with SAML or OAuth2 services.

Figure 1. Login for the NanoCommons Knowledge Base user interface

Home page

You will now be in the Home page (**Figure 2**). On the left you can find a frame providing navigation to the different types of information integrated into the NC-KB.

On the home page screen is a diagram providing a Tiles based overview of the different Information, Actions and Tools currently available. This interface is highly flexible and will change as the project evolves and new information, functions and tools become available and are offered as Transnational Access (TA) services.

The navigation menu (on the left) provides links to content and resources available in the KB as well as the search and management functionalities. The following sections are available:

- Home — links to the most relevant content in the KB (explained below)
- News — links to the NannoCommons project news feed
- My Profile — interface for managing your personal contact data and login information
- Help — link to the help page providing Frequently Asked Questions (FAQ) and guidance for Users, as well as pointing users to the Help Desk (see below)
- Help Desk — link to the electronic issue tracking described in Deliverable report D7.1
- Data Access — links to the available data sets, with search and report functionalities
- Data Upload — interface for uploading data sets into the KB as well as instructions and guidance for users on how to ontologically annotate their datasets for upload into the KB
- Analysis — links to the available analysis pipelines with information on required data formats
- Browse ontologies — links to the available ontologies with browse and search functionalities, again lined to the guidance for users on how to ontologically annotate their datasets for upload into the KB.

Additional menu items will become available as corresponding material is developed, e.g.:

- Training Materials - links to the training materials, video tutorials and other user support offerings developed by NanoCommons partners or tool providers
- Demonstration Case studies - links to the project, industry and regulatory case studies underway or completed and their outcomes.

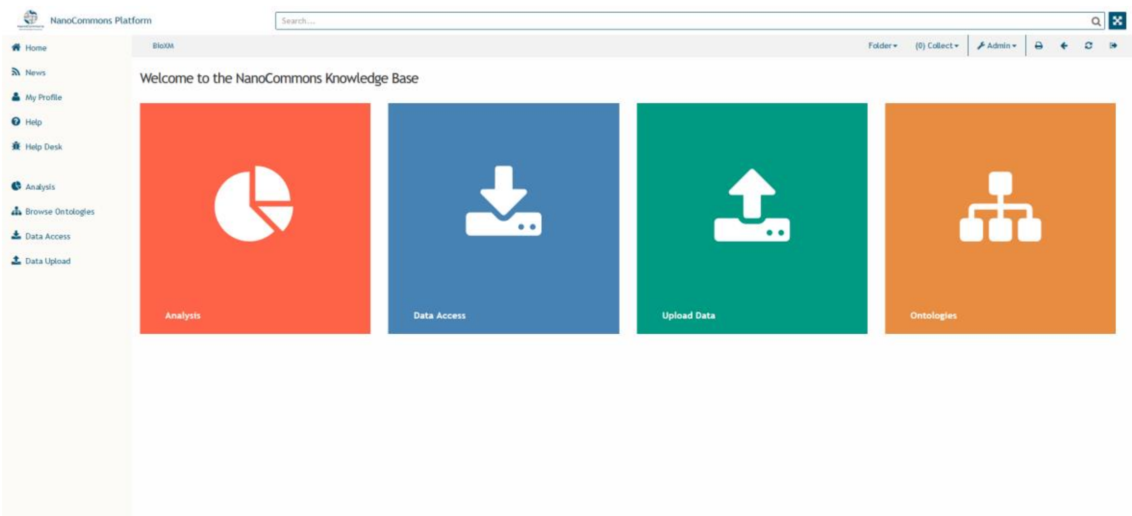


Figure 2. Homepage of the NanoCommons Knowledge Base. The development link is: https://ssl.biomax.de/nanocommons_devel/, but this will be changed to https://ssl.biomax.de/nanocommons/ as we switch the current Development into Production.

Browsing NMs in the knowledge base

Users can browse through the following information related to nanomaterials (NMs) included in the NC-KB:

- NMs General Data (including supplier information, general description, composition information, and European Registry of Materials (ERM) ID number[1])
- NMs Characterisation Data (including ageing and transformations)
- NMs Omics Data
- NMs toxicity / ecotoxicity data
- NMs release, exposure and environmental fate data
- NMs Computational nanodescriptors
- To browse the NMs, click the appropriate link. A table of (currently) 416 NMs will be displayed as the result (**Figure 3**). The table lists the NMs with their respective general information:
- ID (and soon also European Registry ID, a unique and persistent identifier for each NM, including transformed or aged variants, and computational NMs)
- Chemical Elements including any coatings or capping agents
- Basic characterisation data (e.g. size and shape, crystal phase where relevant)
- Samples (batches or lots) including synthesis date (if available) and/or opening date of bottle (for commercial samples) as part of the Provenance information [3]
- Aliquots (where a larger sample was sub-sampled for distribution to project partners)
- Project-specific name where relevant
- Designator (e.g. Producer's identifier or code)
- All names (any synonyms or other short codes used by researchers for this NM)
- Description of physical parameters such as size or form
- Aging reaction (if relevant) [4]
- Storage conditions for the samples to minimise ageing during storage [5]

Clicking on a NM ID will display the "Nanomaterial properties report" for that NM. This report contains the available physicochemical and computational characterisation parameters for each material. These will also be linked to the protocols and methods used for the experimental characterisation or computational models used to calculate the theoretical descriptors that will be included into the NanoCommons protocols repository.

[1] The European Registry of Materials (ERM) was created as an initiative of the NanoCommons project as a simple registry with the purpose to mint material identifiers to be used by research projects throughout the life cycle of their project. ERM identifiers are meant to be used as unique, persistent identifiers to be used in descriptions of experimental designs, in (open) notebooks, in reports, in project milestones and deliverables, and in journal articles.

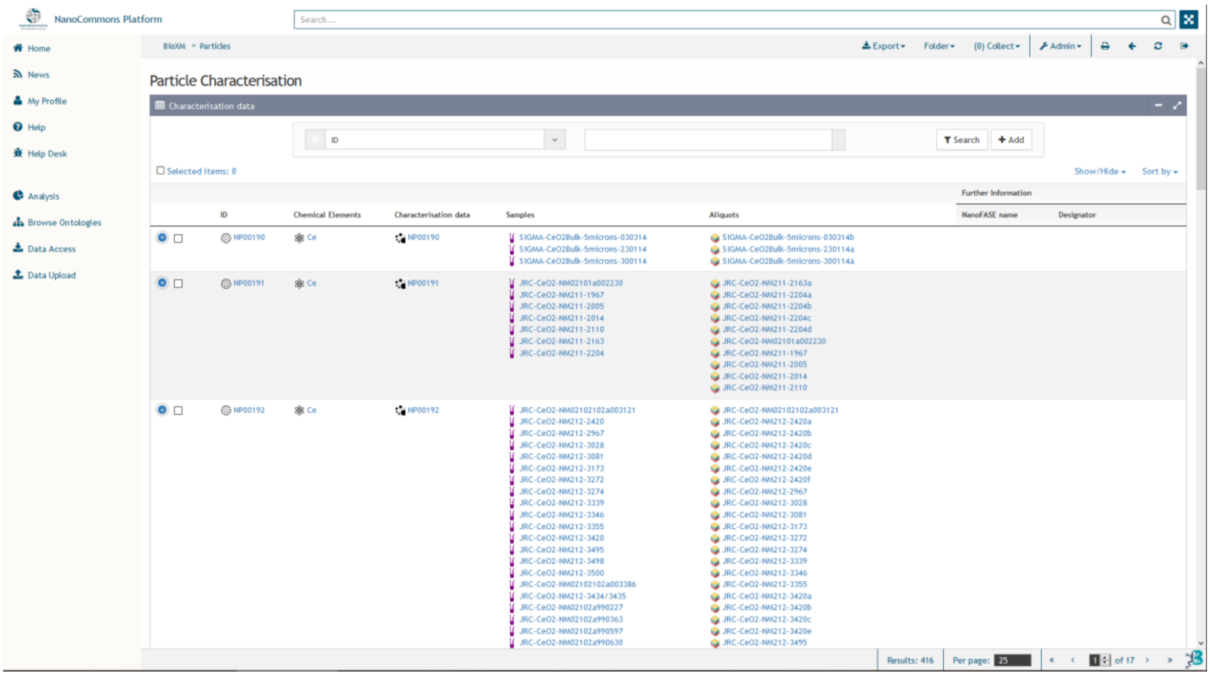


Figure 3. List of NMs currently available in the NanoCommons Knowledge Base, which have come mainly from the NanoMILE and NanoFASE projects to date. A list of literature curated NMs is currently being added, as part of the European Registry of Materials (ERM), developed as part of NanoCommons, and which is currently assigning a persistent and unique Identifier (PId) to each individual NM, including individual batches, aged variants and computationally derived NMs.

Browsing physico-chemical characterization results

To browse NanoCommons Characterization experiments click the "Characterization" link on the "Home" page. The "Particle Characterisation" page, in which the experiments are listed in a table, will be displayed with the characterization data for each NMs (**Figure 4**).

Typically, information provided includes:

- **Size and size distribution** (also called polydispersity index, PDI), indicating by which method data was generated and for electron microscopy, the number of NMs that were analysed.
- **Shape**
- **Crystallinity**
- **Coating composition** and how the coating is attached to the NM
- **Form** in which the NM was supplied (e.g. powder, aqueous dispersion etc.)
- **Dispersion** liquid
- **Redox state** (where relevant)
- **Zeta potential** as a measure of surface charge, which should be accompanied by the pH at which it was measured, and the ionic strength of the measurement liquid.
- **Electromobility**, which is derived from the zeta potential value
- **TEM or STEM** derived size information. If available also images of the NM could be stored and made available to calculate an additional set of computational descriptors, utilising two of the tools integrated into NanoCommons, namely NanoXtract and NanoImage, as

described in Deliverable Report D5.3.

- Energy Band Gap, which is calculated from the UV-Vis spectrum.

However, the availability of these and additional attributes depends on the individual data set, as not all physico-chemical endpoints are relevant to all NMs. For example, redox state is relevant for some metals (e.g. Fe, Ce) but is not relevant for others (e.g. Ti, Si) while some elements can exist in multiple crystal phases (e.g. Ti can exist as anatase, rutile, brookite and amorphous forms) while other elements have only one lattice structure.

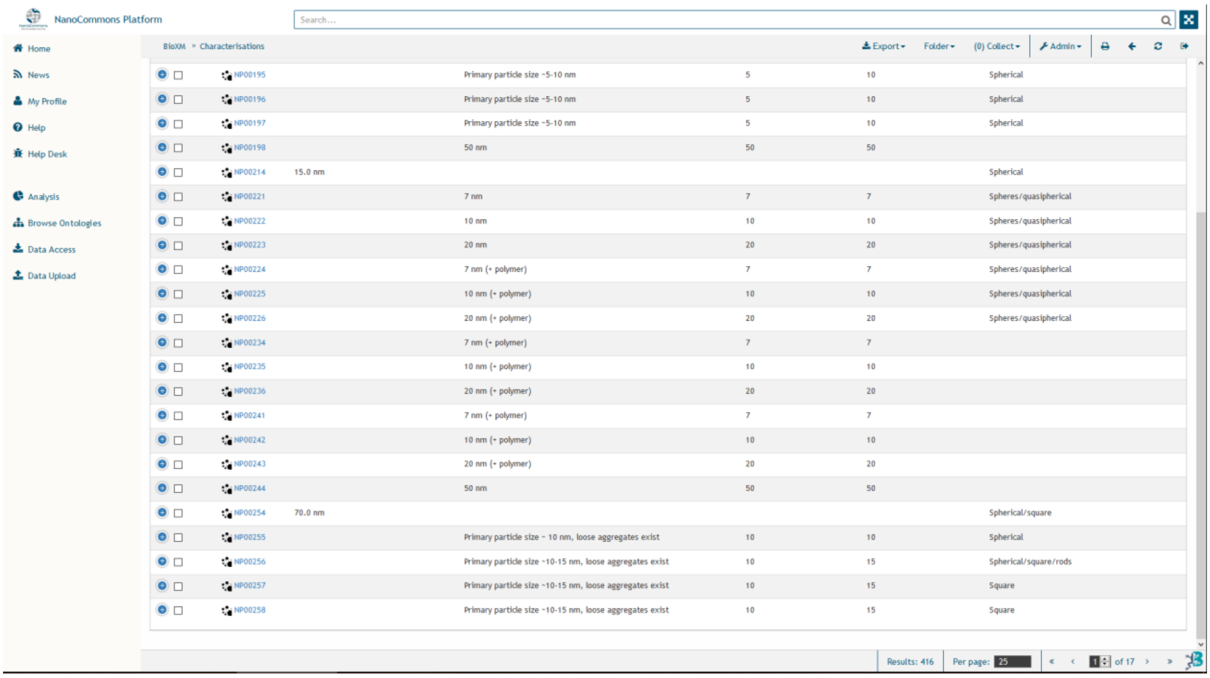


Figure 4. Physico-chemical characterisation available in the NC-KB

Browsing Omics results

To browse NanoCommons Omics experiments click the "Omics" link on the "Home" page. The "RNA-Seq" page in which the experimental results are listed in a table, will be displayed (**Figure 5**).

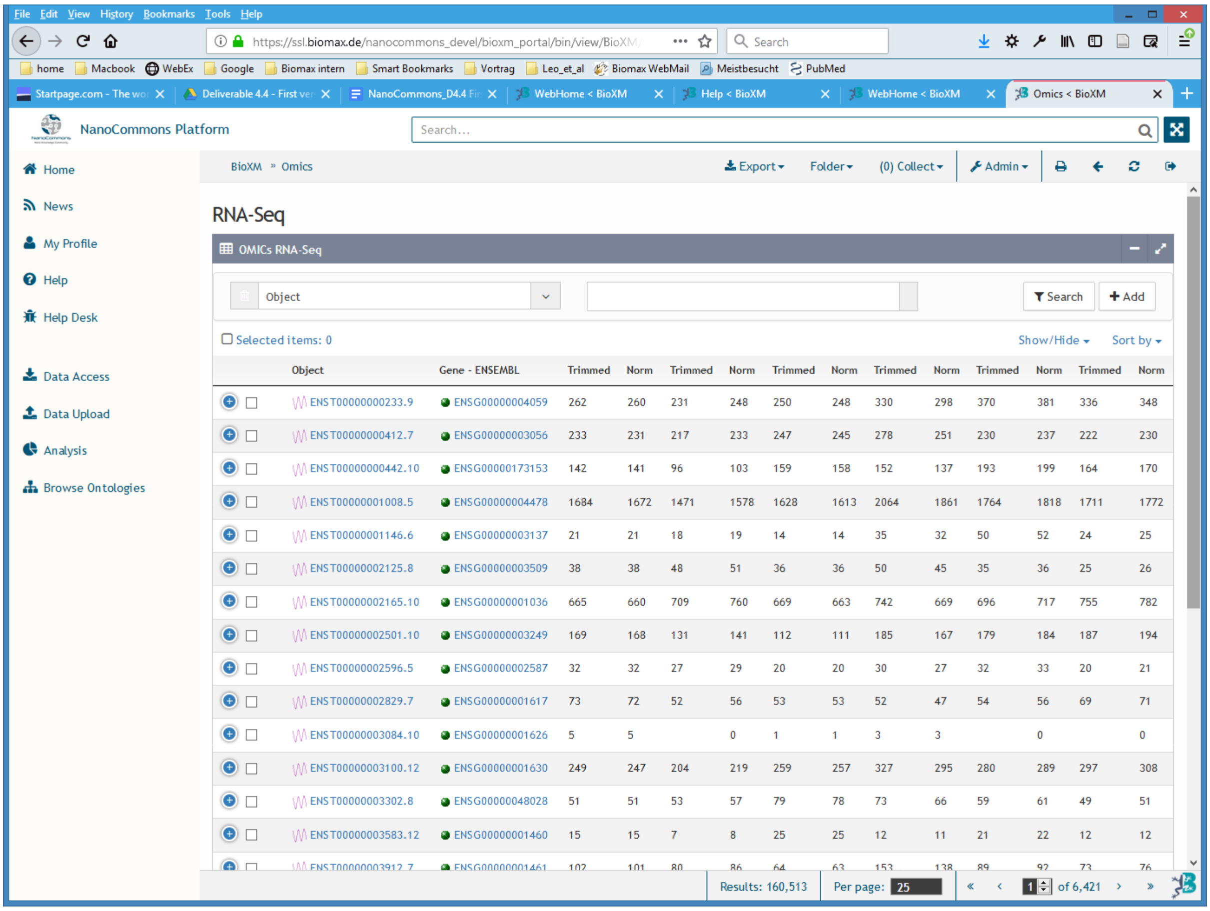


Figure 5. RNA-seq measurements of specific exposure experiments available in the NC-KB

Ontology search tools

To facilitate dataset owners, who are generally experimentalists with limited experience of ontologies or using programming repositories such as GitHub, a simple, user-friendly interface has been implemented that allows users to quickly and easily search for ontological terms and identifiers that match the terms used in their datasets. By annotating the identified ontology terms to their experimental terms, the database then knows exactly where to add each term, and the associated data, into the database to make it easily searchable and retrievable, and to enable combining disparate datasets. Figure 6 shows a screen-shot of the ontology search tool implemented into the NC-KB. A range of relevant ontologies have been incorporated already and, as part of NanoCommons, new terms are being added where needed. More information on the recommended ontologies, the ontology development and the search tools can be found in Deliverable report D4.3.

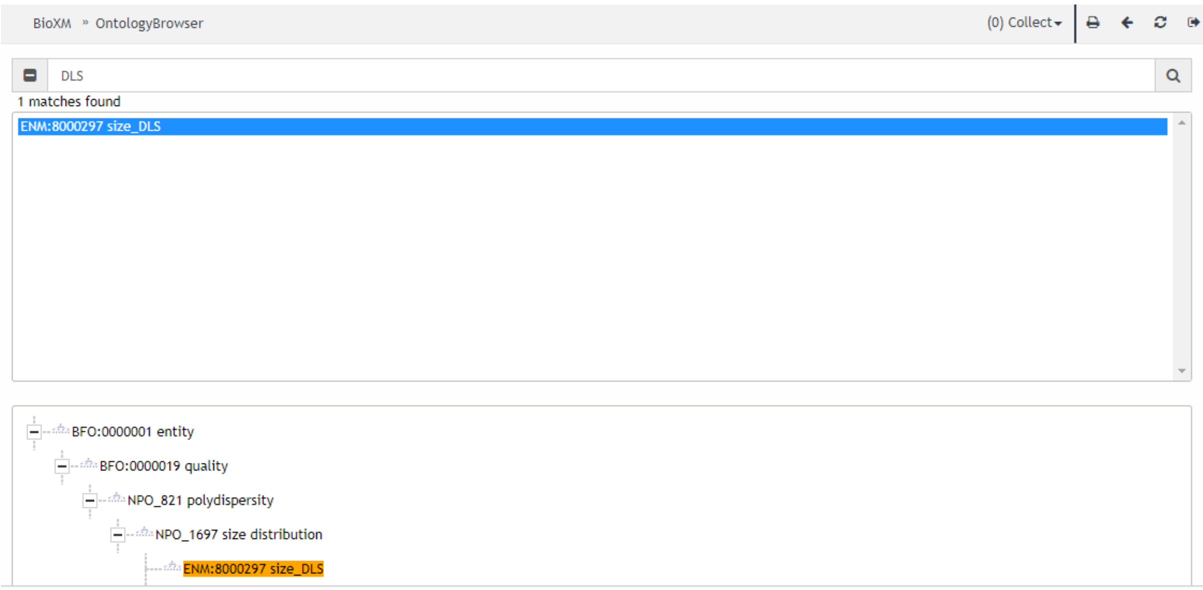


Figure 6: Screenshot of the user-friendly ontology look-up service (BioXM ontology search tool) integrated into the NanoCommons KnowledgeBase (data management module), and integrating the eNanoMapper, the European Materials Modelling Ontology (EMMO), PATO and many other relevant ontologies for NMs, their characterisation, exposure and hazard characterisation and risk assessment.

Data warehouse integration

The NanoCommons data warehouses

As described in detail in deliverable reports D3.3 *Checklist for use in WP8 / WP9 to support integration of Users data into KB*, and D4.5 *Workflow and checklist of key information needed from database/dataset owner in order to facilitate integration into KB,* describing the processes for integration and upload of datasets from individual groups and from large projects / centres, respectively, the goal of NanoCommons is not to develop a single database that sucks in all existing data, but rather to support the nanosafety community to deposit their datasets into the most appropriate existing database where available, and to deposit the rest of the supporting and nanomaterials-specific data into the NanoCommons data warehouse. To achieve this, a number of different APIs have been evaluated for their compatibility with the NC-KB approach, as described in Deliverable Report D4.2 *Initial APIs* which outlined the flexible integration concept and presented examples of how it can be applied to integrate the BioXM and Jaqpot services using their existing APIs, Jupyter notebooks and basic Python commands.

A key aspect of the NanoCommons data warehouse that makes it different from other nano-databases, such as eNanoMapper, is that it incorporates into the database structure and schema the dynamic and context-dependent nature of NMs, whereby the characteristics of a NM at a specific time depend directly on both the NM itself but also the system, and thus the system must be fully described too, and this information becomes part of the data ecosystem allowing direct comparison of the same NM under different biological or environmental conditions. Thus, the NanoCommons data warehouse is adapting the concepts developed in the NanoInformatics Knowledge Commons (NIKC) and improving them to be scalable and automatable across larger scale datasets, and facilitating the utilisation of the extended datasets as input for the various processing, modelling, visualisation and Risk Assessment (RA) tools being developed within NanoCommons.

The NanoMILE and NanoFASE data warehouses

The first version of the approach that has evolved to become the NanoCommons KB was established within the Framework Programme 7 (FP7) project NanoMILE[2] which initially included Biomax as a partner to handle the omics datasets, but which quickly identified a significant gap in how NMs data is captured, and in particular identified a gap in terms of the evolution of NMs in contact with biological and environmental milieu, and the wide range of transformation reactions that NMs undergo. Thus, in NanoMILE, the goal was to develop a system that allowed the characteristics of NMs as received to be linked to the properties of the NMs dispersed in different media, and to the characteristics of the NMs following various ageing and transformation processes, such that we could query the data to identify whether, for example, pristine of transformed properties were more predictive of toxicological impact, and whether transformation increased or decreased the "similarity" of NMs of the same NM produced by different routes or with different initial capping agents, for example, and indeed whether NMs of different compositions aged to similar surface compositions and whether this

[2] http://nanomile.eu-vri.eu/

lead to similar fates and toxicities.

A large volume of data was generated in NanoMILE, spanning NMs characterisation, high content screening analysis in human cells and zebrafish embryos, and toxicity and ecotoxicity studies on a small subset of the total NanoMILE NMs library, which consisted of >100 well characterised NMs. However, the processes for integration of the data were not yet developed while NanoMILE was generating its datasets, and thus while some of the data is integrated into the KB, work is underway to integrate the rest of the data. The physico-chemical characterisation data and the high throughput screening data are integrated (and are also published as in Joossens et al., 2019 [6] and the data deposited in the JRC data repository), as are the omics datasets including some protein corona datasets, and the computational data generated from the Quantitative Structure Activity Relationships (QSAR) models. The data have been uploaded to the Biomax NanoMILE database by the University of Birmingham and are currently being transformed by Biomax's technical team for harmonisation and integration into the NanoCommons Data Warehouse. Additionally, particular data sets are currently under curation by NovaM and UoB in order to make them fit for nanoinformatics modelling, in collaboration with the NanoSolveIT project.

The Horizon 2020 project NanoFASE[3] built upon the initial NanoMILE database, with the NanoFASE data warehouse being version 2.0 of what is now the NanoCommons KB. NanoFASE focussed entirely on characterising (and developing a model to predict) the transformations of NMs in various environmental compartments, including air, water, sediment, soil, during wastewater treatment / in sludge, during incineration and following uptake into organisms and plants. In addition to the individual compartment / species studies, NanoFASE undertook some ambitious mesocosm studies. As a case study for NanoCommons, detailed data capture templates were developed to support the integration of the NanoFASE mesocosm data into the NanoFASE/NanoCommons database, a process which is currently underway very intensively, since NanoFASE is finishing at the end of September 2019. NanoFASE is also focussing on the development of functional assays that can predict key transformations of NMs without needing to run full mesocosms or pilot waste treatment plant scale experiments. The datasets are based on the NIKC template that was developed by NanoCommons partners the Center for the Environmental Implications of Nanotechnology and are described in detail in D3.4 - Guidance document and workflow data templates available for use in WP9. They are Excel based to facilitate data input by experimentalists and enable quick and easy upload. All templates are ontologically mapped to NanoCommons-incorporated ontologies to facilitate the semantic mapping and subsequent data mining, data harvesting and data re-use. Given that data analysis is still underway, the bulk of the NanoFASE datasets are not yet publicly available, but the plan is to release them gradually over the next 24 months as the corresponding papers are published.

The eNanoMapper and related data warehouses

The eNanoMapper project collected various data sets to demonstrate the usability of the ontology and database they were tasked to develop. These data sets have been made available via the data.enanomapper.net instance. Part of this is the CC0-licensed NanoWiki data sets developed by the

[3] http://nanofase.eu/

NanoCommons partner UM, and a protein corona data set for silver NMs developed by NTUA during the eNanoMapper project [7].

Furthermore, during this project, the eNanoMapper partners worked with various other institutes to enable ontological annotation and release of data. Particularly, collaboration between eNanoMapper and NANoREG resulted in ontology annotation for the NANoREG data entry tool and of the NANoREG spreadsheet templates developed by JRC [8]. NANoREG released data at the end of the project using many templates under a permissive CC-BY-NC license. Ingesting this data into the eNanoMapper turned out non-trivial. Because of the number of templates, the diversity of nanosafety research, and the complexity and non-uniformity of the content of the NANoREG templates, data curation on these data files is still ongoing, performed by NanoSafety Cluster projects.

NanoCommons has continued work that started in the last half year of eNanoMapper to allow loading of data in the Resource Description Framework (RDF) format. This offers an alternative way to load data. As an example, a NanoE-Tox data set reported around the potential toxicity of nanomedicine release as Open Data [9] has been made available via the eNanoMapper platform (github.com/egonw/enmrdf/tree/master/NanoE-Tox). Besides the availability of these data sets, various other projects have adopted the eNanoMapper platform to make data available, such as NanoReg2, caLIBRAte, GRACIOUS, and PATROLS. These data sets are currently embargoed and only accessible to a select community (search.data.enanomapper.net) but negotiations are ongoing with the projects to integrate the related data warehouses into the NanoCommons knowledge base once the data becomes publicly available, which will be straightforward due to the common technology and concepts of the public eNanoMapper database and the project-specific warehouses.

The open source eNanoMapper software is under continuous development, with contributions by NanoCommons. First, NanoCommons is developing ELIXIR BioSchemas annotation for content of the database, making the content more Findable (as part of FAIR). Moreover, in collaboration with the BioSchemas team, an extension is under development for Chemical Substances (bioschemas.org/specifications/drafts/ChemicalSubstance/), which can be used for nanomaterials.

A second development is done in collaboration with OpenRiskNet, where the eNanoMapper database infrastructure has been made available as a Docker image and integrated with the OpenRiskNet platform (nanomaterialdb-test.prod.openrisknet.org/ambit/) making it easier for new projects to deploy new instances of this database for managing their data using JRC templates.

Finally, the eNanoMapper comes with an OpenTox-based API, that will automate linking of data into the NC-KB. A recent webinar and developers meeting have been organized to sketch the implementation of such integration.

The ACEnano data warehouses

ACEnano knowledge infrastructure (KI) supports the activities related to data collection and method optimisation in the area of physicochemical characterisation of NMs. The KI provides a central place to access harmonised and standardised methods and data, supporting the implementation of Findable, Accessible, Interoperable and Reusable (FAIR) data principles, the reproducibility and documentation process towards the goal of generating reference resources for NMs RA (**Figure 7**).

With these goals in mind, it fulfils all requirements for becoming a standard resource in the NanoCommons data ecosystem. A public version of the ACEnano data warehouse is being developed and prepared for integration, following the prototype implementation of the eNanoMapper database into NanoCommons. This will help to show the general applicability of the semantic interoperability since a third platform, EdelweissData^{TM}, has then to be supported besides the BioMX and the AMBIT/eNanoMapper platforms.

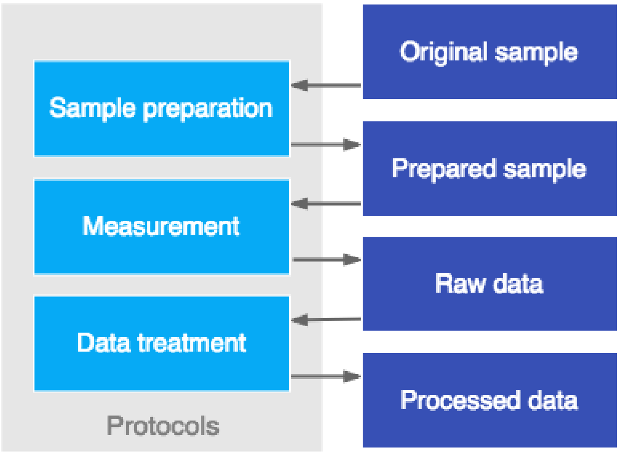


Figure 7. Schematic representation of the strong interlinkage of protocols documented as structured metadata and data stored in the ACEnano data warehouse. Note that the structured metadata protocols approach is also planned to be adopted for NanoCommons as part of the Quality Control / Quality Assurance (QC/QA) approaches, as it minimizes the risk of error in data entry, but yet allows some flexibility of protocols are modified slightly, and allows justification of the modification, such that others can choose to adopt this approach also, and then earlier versions can be retired / withdrawn.

However, it is not only that the data warehouse is coming from another vendor but since ACEnano concentrates, on the one hand on establishing new advanced methods and, on the other hand on inter-laboratory validation of methods, the detailed documentation of metadata describing variations in the experimental setup as part of the data is of absolute necessity and requires new database concepts. Only in this way advances in the methods and improvements in the results obtained from these can be documented and are directly visible to the data user and reasons for reproducibility issues observed in the inter-laboratory studies can be investigated by comparing important experimental parameters stored as metadata. The KI includes instances to accommodate data and protocols. The protocols database facilitates adding, sharing and comparing methods in a questionnaire-like format (**Figure 8**) guiding users through the documentation process from starting material identification to sample preparation, measurement and data processing.

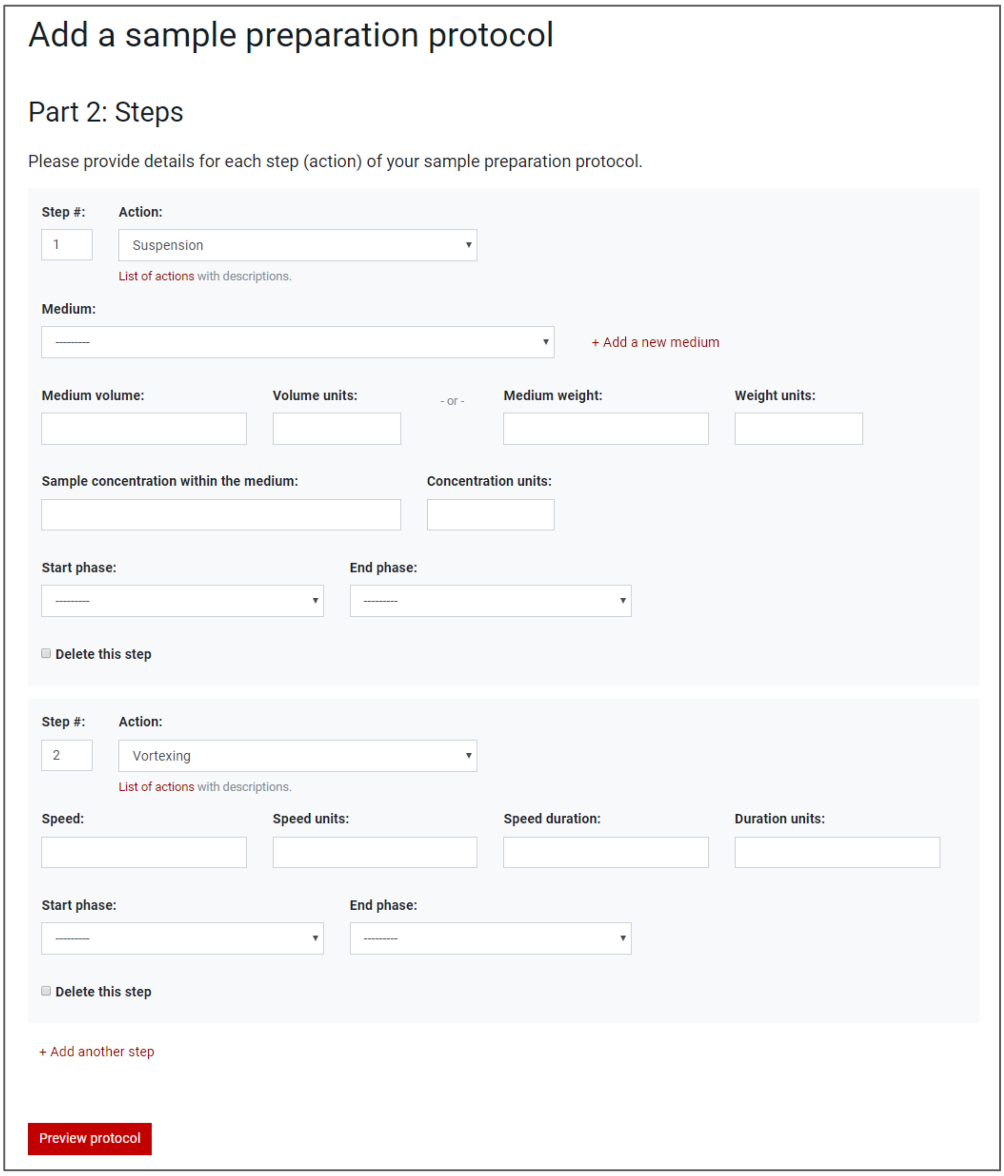


Figure 8. Web interface for creating and filling in information for a protocol in a questionnaire-like format. All actions like “Suspension” or “Vortexing” as well as parameter keys like “Medium” or “Speed” will be semantically annotated to support interoperability with other NanoCommons data sources and facilitate data integration and querying.

During data upload, the user selects and combines protocols according to their experimental procedure and the information in the protocol questionnaires. These are then added as metadata to the data (measurements results) and together stored in the ACEnano data warehouse that offers long-term storage of the results in a reusable, structured and machine-readable format directly linked to the methods applied (**Figure 9**).

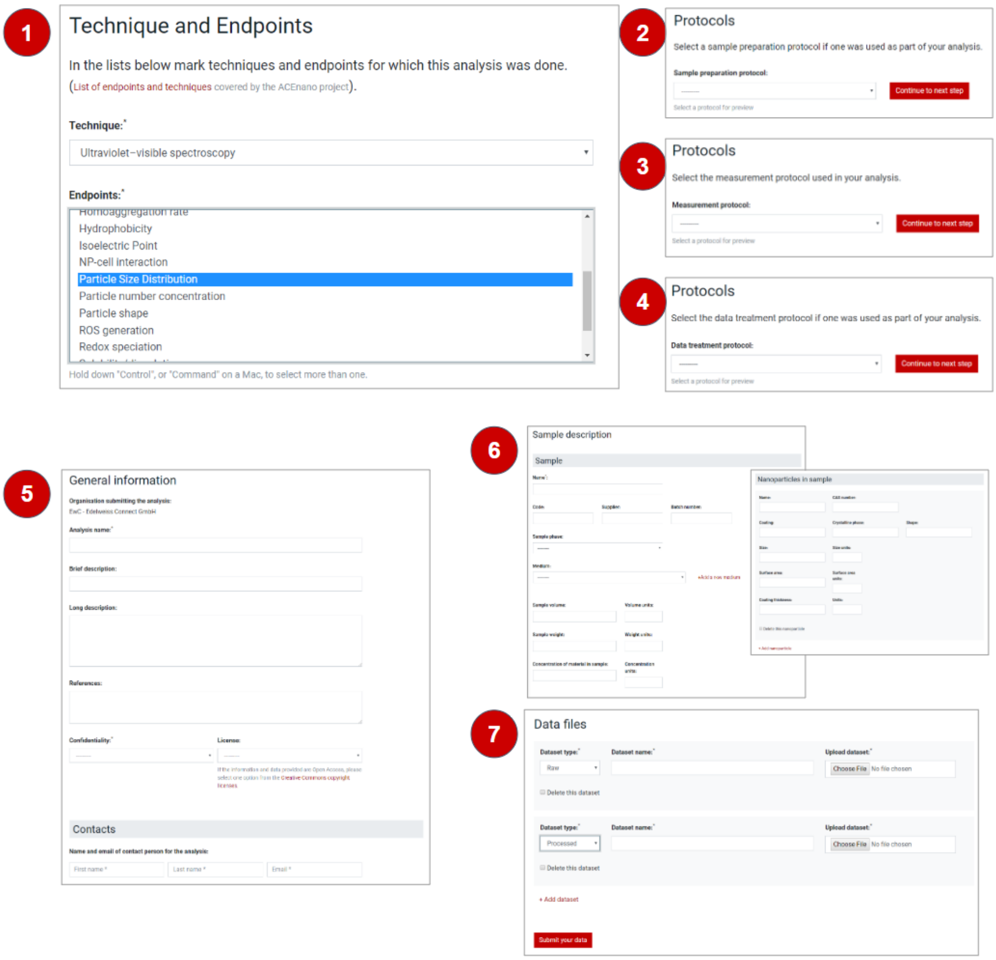


Figure 9. Steps followed by the user in order to create a complete physicochemical characterisation workflow, including the selection of protocols used, description of the sample analysed and data (measurements results) upload.

Even if the final goal is to standardise the methods and protocols and develop harmonised curation templates based on the questionnaires, for the time being the datasets will have the same general format but will include different information for each assay or even for different experiments using the same assay e.g. if different sample preparation methods were applied. Existing and new ontology terms required for the semantic annotation are currently collected (see also the Deliverable D4.3 report on the initial ontology) and concepts to include such variable data schemas in the NanoCommons KB are developed, which is needed to guarantee harmonisation and interoperability with other data sources of the EU NanoSafety Cluster like the eNanoMapper and NanoFASE.

The manual for the ACEnano knowledge infrastructure is available via the NanoCommons GitHub[4], and a training video and step-by-step guide will be prepared in parallel with the integration into NanoCommons for launch as a TA service.

[4] https://github.com/NanoCommons/tutorials/tree/master/ACEnano%20manuals

Data usage via NanoCommons Knowledge Base tools

The overall vision of the NC-KB is to provide interoperability and view based integration of the multitude of existing nano-

data warehouses as described above and thereby offer a single point of access to tools depending on this data. In addition, the integration with data using tools is envisioned as a two-way communication resulting in a platform with integrated data sources and software which presents results from algorithms as yet another characteristic of the analysed nanomaterial, not as substantially different from experimental measurements.

To this end the NC-KB provides a REST Web service API that generically allows to import, query and retrieve all the available information using xml based requests. While this generic API in principle allows collaborators to access any semantically defined item in the NC-KB it requires in-depth knowledge about the implemented semantic model. Therefore, to lower the barrier for collaboration, within NanoCommons we collaborate closely with tool developers to establish pre-defined import templates, queries and data reports that reflect either common or even individual tool needs.

The API based integration of tools with the NanoCommons KB has been described in deliverable reports D4.2, D5.1 and D5.2. Briefly, an omics expression analysis pipeline and a bio-descriptor calculation tool have been integrated, along with various approaches for image analysis, and the GUIDEnano hazard, exposure and RA tool is in the final stage of integration (see Deliverable report D6.1).

As part of the next round of community engagement and user needs analysis, a set of workshops will be organised to develop some key queries that could be developed as tools to interrogate the data in the NC-KB. One key question that springs to mind is whether and how the NC-KB could be queried to answer the fundamental regulatory question of nanoforms, and whether a specific NM is within a group of nanoforms or is a separate one. The European Chemicals Agency (ECHA) have recently released guidance on this, indicating that NMs with different size, surface charge, shape, coatings etc. may be different nanoforms, and that it is up to the registrants to determine and provide evidence as to whether different sizes, shapes, coatings etc. behave the same or differently, and as such constitute one set of nanoforms or multiple sets. “According to Annex VI of REACH: A 'set of similar nanoforms' is a group of nanoforms characterised in accordance with section 2.4 where the clearly defined boundaries in the parameters in the points 2.4.2 to 2.4.5 of the individual nanoforms within the set still allow to conclude that the hazard assessment, exposure assessment and risk assessment of these nanoforms can be performed jointly. A justification shall be provided to demonstrate that a variation within these boundaries does not affect the hazard assessment, exposure assessment and risk assessment of the similar nanoforms in the set. A nanoform can only belong to one set of similar nanoforms.” Thus, NanoCommons is developing a tool, linked to the European Registry of Materials which gives each NM a unique PID, the ECHA template for identifying similarity of NMs, and our models for toxicity and ecotoxicity to enable prediction and testing of boundaries for sets of NMs based on a set of hypotheses regarding the boundaries and the influence of various physico-chemical parameters, coupled with mechanistic models for exposure, hazard and RA.

NanoCommons Knowledge Infrastructure catalogues

The **NanoCommons Knowledge infrastructure** (https://infrastructure.nanocommons.eu/) has currently three distinguished but interlinked sections:

- **Services**: The NanoCommons e-infrastructure aims to integrate and further develop existing state-of-the-art tools and to develop those that are needed to fill in the experimental, computational and beyond the needs of the nanosafety community. The services are covering several areas, like data storage and online accessibility, data visualisation and predictive toxicity, data processing and analysis or experimental workflow design & implementation.
- **Library**: This page contains resources and training materials to support NanoCommons users in getting familiar with the services and tools available in the infrastructure. On top of tutorials and video demonstrations, you will also find information on our publications (e.g. peer-review articles, presentations, posters) that may help you further in learning about NanoCommons concepts and implementations.
- **Events**: List of conferences and other events like workshops, hackathons, trainings or webinars organised or attended by NanoCommons members, including the links to the relevant materials generated.

As described in Deliverable 5.1. one of the central entry points to the NanoCommons KB is the NanoCommons Service Catalogue, directly accessible from the NanoCommons website. The catalogue is based on the technology developed in the e-infrastructure project OpenRiskNet[5] that has been specifically adapted to the needs of NanoCommons (**Figure 10**).

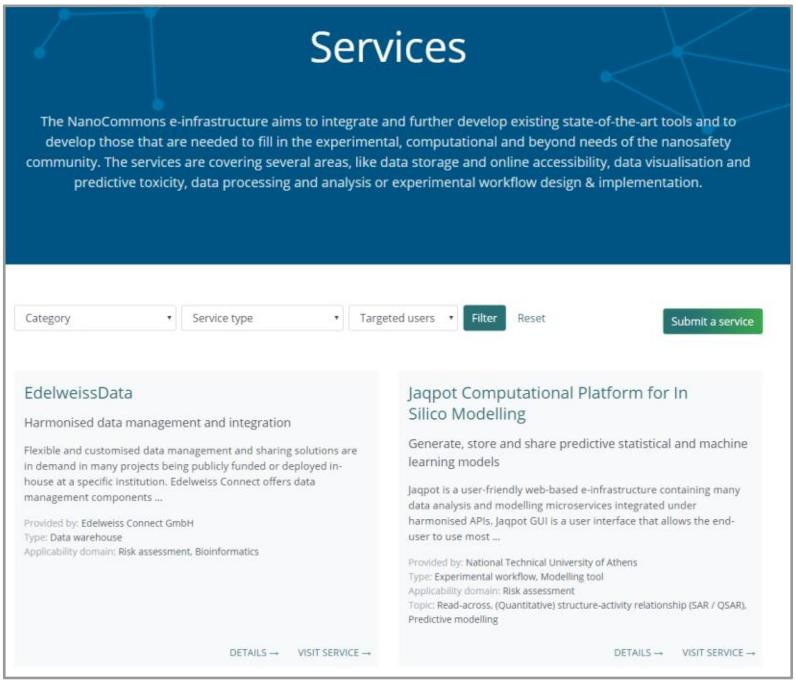


Figure 10. Public interface of the NanoCommons catalogue of services

[5] https://openrisknet.org/

The services listed are directly linked to the NanoCommons Library (collection of training materials, publications and other resources for users) and a page dedicated to relevant NanoCommons Events.

Services description

The catalogue provides a detailed description of the nanoinformatics services currently offered by NanoCommons for TA and/or remote access, and provides direct links to the service environment, their APIs and to all related support resources. The catalogue supports the users in filtering the information on services offered and the corresponding tools based on predefined descriptors, i.e.:

- Category of services
- Service type
- Targeted users
- Data inputs required

Additional filters can be implemented using the structure of the catalogue (**Table 1** and **Figure 11**).

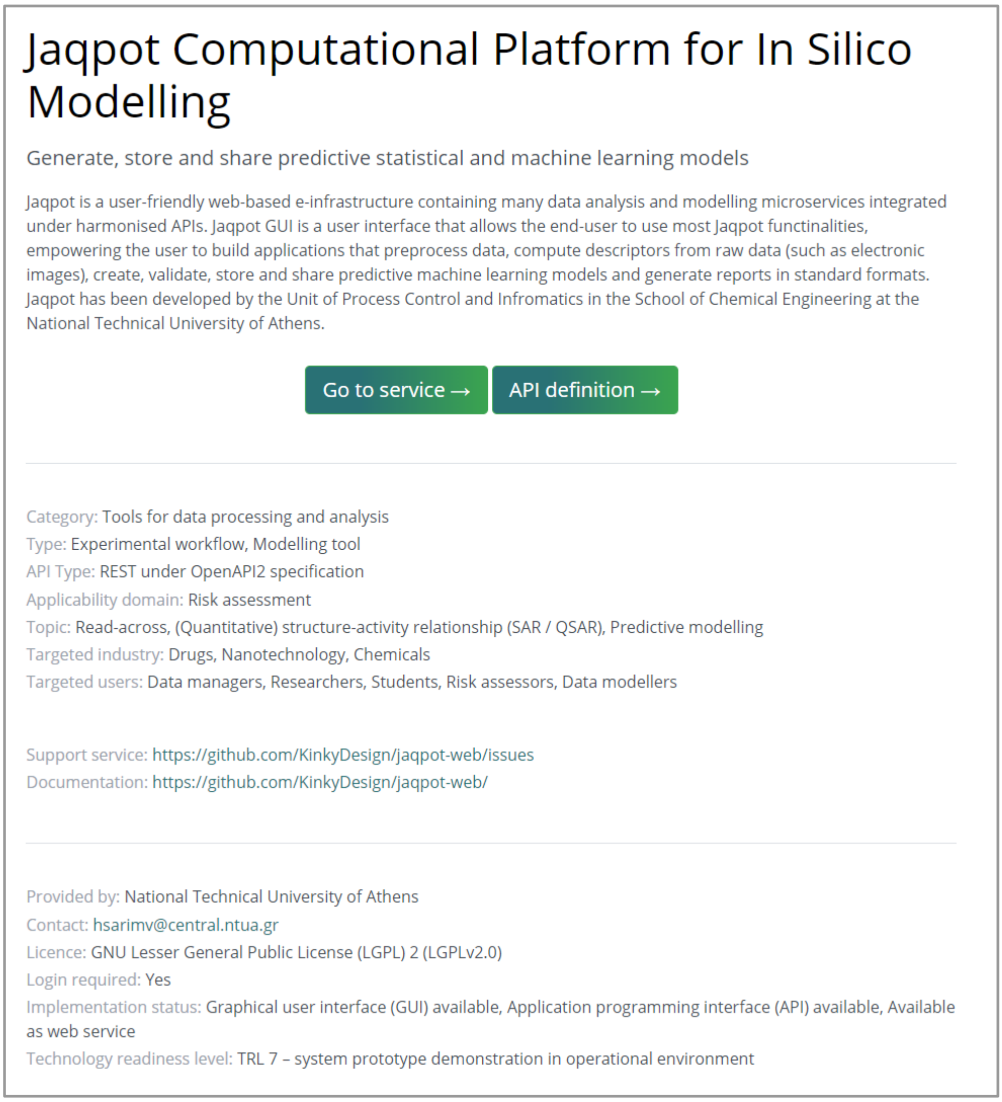
Deliverable report D5.1 also summarises the full range of services that have been implemented to date, covering each of the four categories of TA services offered by NanoCommons, as follows:

Table 1. Sections and descriptors currently used for the description NanoCommons services

Service identification
Name
URL
API URL
API Type
Provider name
Provider contact
Provider organisation
Service description
Tagline: brief free text
Description: brief free text
Category (*multiple selection*): ● Tools for data storage and online accessibility ● Tools for data visualisation and predictive toxicity ● Tools for data processing and analysis ● Tools for experimental workflow design & implementation
Service type (*multiple selection*): ● Knowledge base ● Data warehouse

• Modelling tool • Semantic annotation tool • Data curation tool • Image analysis tool • Protocols and methods repository • Experimental workflow • Electronic Laboratory Notebook (ELN)
Implementation status (*multiple selection*): • Graphical user interface (GUI) available • Containerised • Application programming interface (API) available • Available as web service
Technology readiness level: (*single selection*) from TRL-1 to TRL-9
Applicability domain (*multiple selection*): • Hazard assessment • Risk assessment • Risk characterisation • Bioinformatics • Exposure assessment • Ontologies
Topic (*multiple selection*): • Read-across • (Quantitative) structure-activity relationship (SAR / QSAR) • Protein and small molecule corona analysis • Information extraction • Identifier mapping • Kinetics / biokinetics • Predictive modelling • Omics data analysis • Physicochemical characterisation of nanomaterials • Toxicology • Ecotoxicology
Targeted industry (*multiple selection*): • Cosmetics • Drugs • Nanotechnology • Chemicals • Other consumer products • Food and feed • Textiles • Constructions • Automotives
Targeted users (*multiple selection*): • General public • Regulators

• Data managers • Software developers • Researchers • Students • Risk assessors • Policy makers • Data modellers
Licence type (*single selection*): • Open source • Proprietary software
Licence: various options available
Login required
Training and user support
User support service
User support contact
Documentation center
References



Jaqpot Computational Platform for In Silico Modelling

Generate, store and share predictive statistical and machine learning models

Jaqpot is a user-friendly web-based e-infrastructure containing many data analysis and modelling microservices integrated under harmonised APIs. Jaqpot GUI is a user interface that allows the end-user to use most Jaqpot functinalities, empowering the user to build applications that preprocess data, compute descriptors from raw data (such as electronic images), create, validate, store and share predictive machine learning models and generate reports in standard formats. Jaqpot has been developed by the Unit of Process Control and Infromatics in the School of Chemical Engineering at the National Technical University of Athens.

Go to service → API definition →

Category: Tools for data processing and analysis
Type: Experimental workflow, Modelling tool
API Type: REST under OpenAPI2 specification
Applicability domain: Risk assessment
Topic: Read-across, (Quantitative) structure-activity relationship (SAR / QSAR), Predictive modelling
Targeted industry: Drugs, Nanotechnology, Chemicals
Targeted users: Data managers, Researchers, Students, Risk assessors, Data modellers

Support service: https://github.com/KinkyDesign/jaqpot-web/issues
Documentation: https://github.com/KinkyDesign/jaqpot-web/

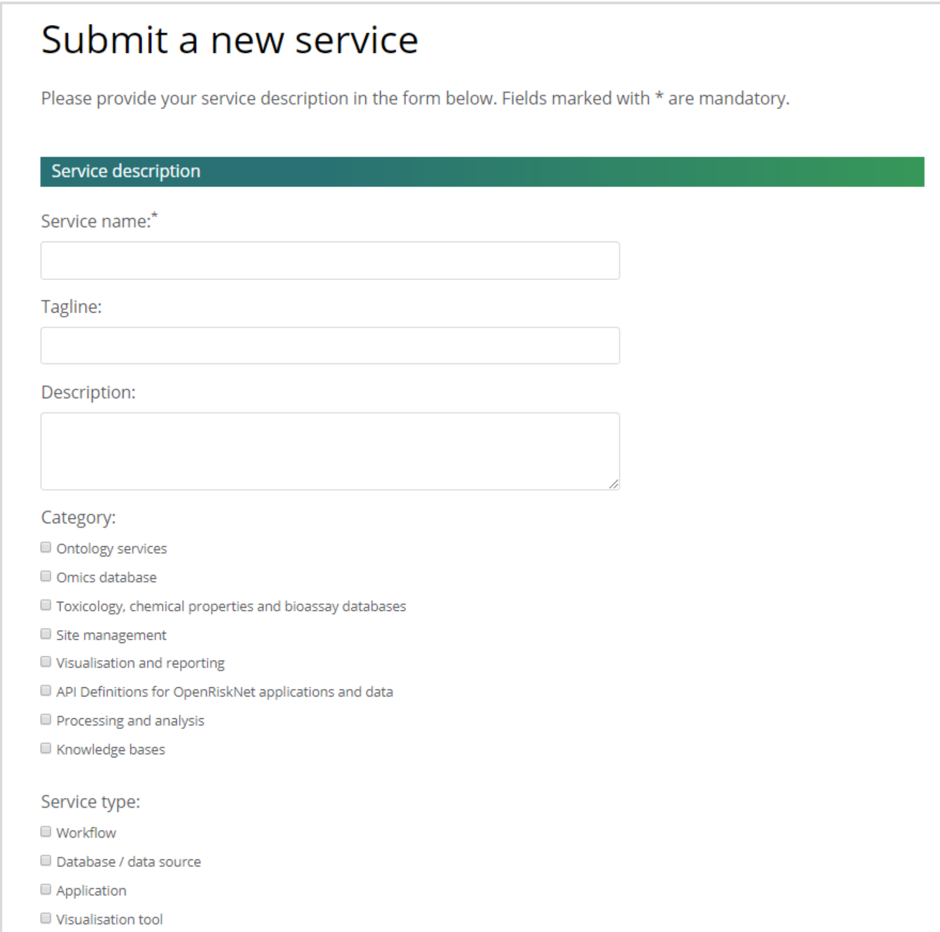
Provided by: National Technical University of Athens
Contact: hsarimv@central.ntua.gr
Licence: GNU Lesser General Public License (LGPL) 2 (LGPLv2.0)
Login required: Yes
Implementation status: Graphical user interface (GUI) available, Application programming interface (API) available, Available as web service
Technology readiness level: TRL 7 – system prototype demonstration in operational environment

Figure 11. Example of a nanoinformatics service included and described in the NanoCommons services catalogue - the Jaqpot workflow for generation of predictive statistical and machine learning models, developed by NTUA.

Addition of new services

NanoCommons partners and stakeholders have the possibility to add and describe new services and tools. This is done by using the function “Submit a new service”[6] (**Figure 12**), which allows the service providers to describe in detail their tools using a predefined online form that contains all descriptors shown in **Table 1** above. In addition, the submitter needs to agree with the ‘Privacy Policy’ required for the processing, approval and publishing of the data provided to NanoCommons.

Once submitted, the information is reviewed by the catalogue administrators and marked as approved, therefore ready to be published online and visible to the public. If more information is required, the service provider is contacted in order to provide all necessary details before making it available in the catalogue.


Submit a new service

Please provide your service description in the form below. Fields marked with * are mandatory.

Service description

Service name:*

Tagline:

Description:

Category:

- Ontology services
- Omics database
- Toxicology, chemical properties and bioassay databases
- Site management
- Visualisation and reporting
- API Definitions for OpenRiskNet applications and data
- Processing and analysis
- Knowledge bases

Service type:

- Workflow
- Database / data source
- Application
- Visualisation tool

Figure 12. Online form used for addition of new services to the NanoCommons catalogue of services.

[6] https://openrisknet.org/nanocommons/add/

Conclusions

The NanoCommons knowledge base user interface, the NanoCommons data warehouse and the NanoCommons service catalogue are now fully operational as the first three major components of the knowledge infrastructure.

The user interface is the main entry point for users to find information on specific nanomaterials and search, browse and access specific datasets from the linked data resources.

The data warehouse is ready to be offered as a service to projects and individual researchers to store, manage and share their data based on the FAIR principles. Data sets from NanoMILE, NanoFASE and the NanoInformatics Knowledge Commons (NIKC) have been uploaded and work is ongoing to achieve the complete coverage of all data from these projects.

The service catalogue provides information of all the integrated NanoCommons data and software offerings. It helps the user to find resources relevant to their work for direct integration in their work or which could be explored and optimized to their needs as part of a Transnational Access offering.

References

1. Losko S, Heumann K. Semantic data integration and knowledge management to represent biological network associations. Methods Mol Biol. 2009;563: 241–258.

2. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, et al. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst Biol. 2011;5: 38.

3. Baer DR, Munusamy P, Thrall BD. Provenance information as a tool for addressing engineered nanoparticle reproducibility challenges. Biointerphases. 2016;11: 04B401.

4. Baer DR. The Chameleon Effect: Characterization Challenges Due to the Variability of Nanoparticles and Their Surfaces. Front Chem. 2018;6: 145.

5. Izak-Nau E, Huk A, Reidy B, Uggerud H, Vadset M, Eiden S, et al. Impact of storage conditions and storage time on silver nanoparticles' physicochemical properties and implications for their biological effects. RSC Adv. 2015;5: 84172–84185.

6. Joossens E, Macko P, Palosaari T, Gerloff K, Ojea-Jiménez I, Gilliland D, et al. A high throughput imaging database of toxicological effects of nanomaterials tested on HepaRG cells. Sci Data. 2019;6: 46.

7. Jeliazkova N, Chomenidis C, Doganis P, Fadeel B, Grafström R, Hardy B, et al. The eNanoMapper database for nanomaterial safety information. Beilstein J Nanotechnol. 2015;6: 1609–1634.

8. JRC:Joint Research Centre. NANoREG data logging templates for the environmental, health and safety assessment of nanomaterials. Publications Office of the European Union; 2017.

9. 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.