



MESOCOSM: A mesocosm database management system for environmental nanosafety

Ali Ayadi, Jérôme Rose, Camille de Garidel-Thoron, Christine Hendren, Mark R Wiesner, Melanie Auffan

► To cite this version:

Ali Ayadi, Jérôme Rose, Camille de Garidel-Thoron, Christine Hendren, Mark R Wiesner, et al.. MESOCOSM: A mesocosm database management system for environmental nanosafety. NanoImpact, 2021, 21, pp.100288. 10.1016/j.impact.2020.100288 . hal-03150076

HAL Id: hal-03150076

<https://cnrs.hal.science/hal-03150076>

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MESOCOSM: a mesocosm database management system for environmental nanosafety

Ali Ayadi^{a,1}, Jérôme Rose^{a,b}, Camille de Garidel-Thoron^a, Christine Hendren^b, Mark R. Wiesner^b,
Mélanie Auffan^{a,b}

^a*CEREGE, CNRS, Aix Marseille Univ, IRD, INRAE, Coll France, Aix-en-Provence, France*

^b*Duke university, Civil and Environmental Engineering, Durham, USA*

Abstract

Engineered nanomaterials (ENMs) are intentionally designed and produced by humans to revolutionize the manufacturing sector, such as electronic goods, paints, tires, clothes, cosmetic products, and biomedicine. With the spread of these ENMs in our daily lives, scientific research have generated a huge amount of data related to their potential impacts on human and environment health. To date, these data are gathered in databases mainly focused on the (eco)toxicity and occupational exposure to ENMs. These databases are therefore not suitable to build well-informed environmental exposure scenarios covering the life cycle of ENMs. In this paper, we report the construction of one of the first centralized mesocosm database management system for environmental nanosafety (called MESOCOSM) containing experimental data collected from mesocosm experiments suited for understanding and quantifying both the environmental hazard and exposure. The database, which is publicly available through <https://aliayadi.github.io/MESOCOSM-database/>, contains 5200 entities covering tens of unique experiments investigating *Ag*, *CeO₂*, *CuO*, *TiO₂*-based ENMs as well as nano-enabled products. These entities are divided into different groups i.e. physicochemical properties of ENMs, environmental, exposure and hazard endpoints, and other general information about the mesocosm testing, resulting in more than forty parameters in the database. The MESOCOSM database is equipped with a powerful application, consisting of a graphical user interface (GUI), allowing users to manage and search data using complex queries without relying on programmers. MESOCOSM aims to predict and explain ENMs behavior and fate in different ecosystems as well as their potential impacts on the environment at different stages of the nanoproducts lifecycle. MESOCOSM is expected to benefit the nanosafety community by providing a continuous source of critical information and additional characterization factors for predicting ENMs interactions with the environment and their risks.

Keywords: Mesocosm testing, Database management system, Environmental exposure and hazards, Risk assessment, FAIR data.

2010 MSC: 00-01, 99-00

1. Introduction

A growing number of companies are manufacturing or using nanomaterials [1]. These engineered nanomaterials (ENMs) play an important role in the innovation of many industrial sectors such as construction (self-cleaning concrete), cosmetics (sunscreens), textiles (antibacterial clothing), health (improving medical imaging), paint (scratch resistant paint), etc. [2, 3]. However, the large specific surface area and the enhanced surface reactivities at the nanoscale have raised several questions these past 20 years regarding their potential environmental risk (Exposure \times Hazard) if released in the Environment. Several groups around the world have addressed this issue and generated data based on nanoecotoxicology approaches [1] focusing on a single species or strain of model organisms. Such experiments are needed to gain knowledge on the mechanisms and thresholds of toxicity (Hazard), but they separate the organism from their environment (relevant Exposure) and community of organisms (i.e., the ecosystem), in which the environmental risks of ENMs ultimately need to be understood. For a robust characterization of the environmental risk, the complex ecosystem-level interplay between the organisms, their environment, and the ENMs needs to be taken into account including but not limited to: aggregation state and sorption of (in)organic substances [2, 3], oxidation-reduction potential, as well as ecological factors such as interacting organisms (e.g. microorganisms, (in)vertebrates, plants, fungi), trophic levels present (e.g. primary producer, primary consumer, secondary consumer) and trophic transfer potential [4, 5]. Particularly well-suited experimental devices to engage such integrated and multidisciplinary approach are mesocosms [6, 7, 8]. In the scope of ENM studies, mesocosms have been defined as indoor or outdoor experimental systems containing a portion of the natural environment which is (i) self-sustaining once set up and acclimatizes without any additional input of nutrients or resources, and (ii) that allows monitoring all (or the maximum of) input and output parameters to draw a real-life mass balance [9].

For years, the SERENADE, for Safe(r) and Ecodesign Research and Education applied to NANomaterial Development, consortium in collaboration with the Center for the Environmental Implications of Nanotechnology (CEINT) at Duke University, has been continuously conducting mesocosm testing to address environmental nanosafety throughout the lifecycle of nanoproducts, from the early stages of production to the end of life [10]. However, the experimental data produced were stored in non-standardized forms limiting the accessibility, interoperability, and reusability of the generated data for further applications. To meet this need, the development of a centralized mesocosm database management system for environmental nanosafety that will allow to collect, standardize, and store all the exposure and hazard experimental data obtained in mesocosms was required [5, 9].

Consequently, the mesocosm database management system for environmental nanosafety (MESOCOSM) presented in this paper, will be one of the first free access database systems providing and managing robust environmental exposure and hazard data. The proposed system aims to achieve a highly organized relational database along with an appropriate software application, consisting of a graphical user interface (GUI), that facilitate processing and access to that data. The MESOCOSM database provides 5200 unique experimental data on tens of unique mesocosm experiments based on *Ag*, *CeO₂*, *CuO*, *TiO₂*-based ENMs as well as nano-enabled products [11, 12, 13] with different shape, size, surface coatings and matrix. These data were collected from mesocosm testing to address environmental nanosafety throughout the lifecycle of nanoproducts, from the early stages of production to the end of life. MESOCOSM will be an added-value to model and forecast the behavior and fate of ENMs in the environment, as well as to assess their potential environmental risks.

2. Overview of nanomaterial-related databases

Over the last decades, experimental studies on (eco)toxicologically, behavior and fate of ENMs have significantly increased. This evolution comes along an increase in the number of databases (non-exhaustively listed below), knowledge base [14, 15], web-based libraries [16], or also registry repositories [17, 18, 19] with the financial support of the EU and US international organizations. An overview of ENM-related databases is provided in Table 1.

Several of these databases are not specific to ENMs. Among others, the National Toxicology Program database (NTP) contain modern toxicology and molecular biology data to identify substances (ENMs and others) in the environment that may affect human health [20]. The ECETOC Aquatic Toxicity database (EAT) is an ecotoxicological database regarding the toxicity of diverse substances to aquatic species in fresh and saline waters [21]. The EAT database gives access to the Human Exposure Assessment Tools Database (heatDB) which collects all publicly available sources of human exposure data. The National Institute for Occupational Safety and Health database (NIOSH) provide databases focusing on occupational toxicology information (ENMs and others) [22]. One last example, is the Ecotoxicological Risk Assessment database (ERA) that includes the concentration ranges for various human-use drug residues, their potential effects for aquatic organisms, and the temporal variations of the ecotoxicological risks [23].

Of course, with the enhanced research and development activities around Nanosciences and Nanotechnologies, ENM-specific databases were developed. Most of them focus on toxicity and/or ecotoxicity data. For instance, the NanoE-Tox is a database on the nanoecotoxicity of eight ENMs that gathers data from more than 220 scientific articles [24]. The Nano Environmental Health and Safety database (NanoEHS) concerns the physico-chemical properties of ENMs, their biocompatibility and potential toxicity to human health [25]. The NanoPUZZLES dataset gathers toxicity and cytotoxicity data. NanoEHS, NanoPUZZLES are used to predict the relationships among the structure, properties, molecular interactions, and toxicity of ENMs [26, 27]. The NanoMILE database aims to understand the mechanisms of ENMs interactions with living systems and the environment [28]. Both NanoPUZZLES and NanoMILE used the ISA-TAB-Nano template as an ENMs data exchange standard [29]. Other examples are the Nano Exposure and Contextual Information database (NECID) that collects exposure measurements of nano objects and their agglomerates and aggregates [30] in an occupational exposure perspective, or also the GUIDEnano database about the human health risks of nano-enabled products all along the product life cycle [31]. Finally, one of the larger European database is eNanoMapper that integrates generic nanosafety data from the scientific literature. eNanoMapper includes functionalities for data protection and sharing, data quality assurance, and search interfaces for different usages [32]. eNanoMapper is part of the computational infrastructure for toxicological data management of ENMs.

Several of these ENM-databases are domain specific and are addressed to specific communities. For instance, the NanoMiner database is designed for toxicogenomic researchers and contains more than 400 human transcriptome samples exposed to various types of ENMs [18]. The CaNanoLab database is intended for the biomedical community who wants to use of nanotechnologies in biomedicine [33]. The NanoToxDB database provides information to biologists on ENMs toxicity to a specific aquatic invertebrate, the *Daphnia magna* [34]. The InterNano database is dedicated to the nanomanufacturing community [35]. It shares standard nanomanufacturing processes and information on the properties of ENMs across sectors [36] to facilitate the commercial development of nanotechnology.

Regarding the environmental exposure to ENMs, only two databases specifically address this domain: the NanoInformatics Knowledge Commons database (NIKC) and the ECOTOXicology knowledgebase (ECOTOX). The NIKC, developed by the Center for Environmental Implications of NanoTechnology (CEINT), provides a critical mass of experimental data generated in the CEINT center and related to environmental exposure and hazard data. The ECOTOX is an open access repository including single chemical toxicity data for aquatic life, terrestrial plants, and wildlife [37]. Developed by the Center for Computational Toxicology and Exposure (CCTE) and the Great Lakes Toxicology Ecology Division (GLTED) which are both components of the USEPA Office of Research and Development, the ECOTOX database incorporates three independent databases of ecotoxicity data from various literature sources the AQUIRE, PHYTOTOX, and TERRETOX focusing on aquatic life, terrestrial plants, and terrestrial wildlife, respectively.

Finally, few of these databases act as networking tools as the NanoWerk Nanomaterial database which provides information about nano conferences, nano events, nano companies, laboratories, and nanotechnology applications [38], as well as the nanoHub database which provides information about available resources (from video presentations to publications to simulation tools) [39].

Most of these databases were based on data extracted from published literature. Interestingly, few of them combine the database with modeling tools to go further in the assessment and forecasting of the risks for human and environmental health. As an example, the Nanomaterial-Biological Interactions Knowledge

Base (NBI Knowledge Base) is designed to interpret nanomaterial exposure effects in biological systems [40]. It acts as a depository for data but is also equipped with data mining tools to predict the biological interactions of ENMs. The Online Chemical Modeling Environment database (OCHEM) provides experimental measurements and properties associated to chemical compounds [41]. It also incorporates a modeling framework to create a predictive model of biological and physicochemical properties of chemical compounds [42]. One last example is the Modeling Assays Platform for hazard ranking of engineered nanoparticles database (MOD-ENP-TOX) and the Nanomaterials Experiment-based Predictor of Environmental Risk and Toxicity (NanoExPert) database that contains information about nanoparticles properties and their interaction with biological systems [23, 43, 44]. Both of them act as tools to predict the toxicity and ecotoxicity of ENMs.

This literature survey of the existing databases on ENMs highlights that most of them focused either on the properties of ENMs, their hazards (toxicity, ecotoxicity, bioavailability) or the occupational exposure. However, there is an obvious lack of databases about environmental exposure to ENMs. Gathering data about the exposure and hazard of ENMs within ecosystems and taking into account different lifecycle stages of ENMs-based products, the MESOCOSM database will fill this gap. MESOCOSM will be integrated with the NIKC [50] and the eNanoMapper [32] databases. It will also be integrated into the search engine of the international NanoInformaTIX database. NanoInformaTIX is a comprehensive database integrating several relevant databases from EU/US projects supporting various aspects of ENMs nanosafety (materials, hazard, environmental release, occupational exposure, omics, bio-nano interactions, pharmacokinetics, . . .), to improve predictive (eco)toxicity assessment of ENMs during their whole life cycle. By itself, the development of the MESOCOSM database will be an added-value in the field and its integration with other nano-database will provide the needed tool for sustainable innovation pathways of nanotechnologies. In particular, the quality of the data provided by the MESOCOSM database may support manufacturers and researchers to consider the behavior of ENMs under different environmental conditions and during their entire life-cycle in order to provide a complete view of their potential risks on the environment.

3. MESOCOSM database description

3.1. MESOCOSM database challenges

The database is available to all on GitHub repository. Scientists and industries can visualize the totality or a part of the dataset, download the SQL database file and manipulate it with any database management system (Oracle, Postgres, MySQL, DB2, . . .), remotely interact with the database via an application program interface (API), or also download it with its application (graphical user interface) for local usage (on the local computer or server).

As described in Figure 2, the application provides dialog boxes displayed on the frontend layer to interact with users, allowing either (i) to notify the user that the input (insertion or modification) has been successfully done, (ii) to confront the user with a choice, or also (iii) to ask the user to fill some missing fields, and (iv) to ask the user to fulfill a request form or select search criteria to retrieve information or refine the request, respectively. Through this graphical interface, users can interact with the MESOCOSM database more flexibly and intuitively.

It is also expected that this application could encourage users to use the MESOCOSM database and make it evolve. There are currently 5200 entities in the MESOCOSM database related to tens of unique experiments investigating CeO_2 , CuO , TiO_2 -based ENMs, and nano-enabled products. But we hope that with collaborative efforts, this number of entities could quickly evolve. Indeed, the database website invites the large nanotechnology and nanoscience communities to submit additional data that could contribute to the enrichment of the MESOCOSM database, as well as to address suggestions to improve the database.

The MESOCOSM database provides experimental data of interest to a wide range of users, including manufacturers, researchers, and government agencies. It allows them to consult, update, export data on environmental exposure to ENMs. With MESOCOSM, results obtained by international research groups can be shared to the scientific community. This will encourage researchers to avoid repeating experiments that have already been done, reject some research hypotheses and favor others. Only processed data, textual, and already published are stored in the MESOCOSM database. Raw, non-numerical, unpublished data are not

Table 1: A summary of nanomaterial-related databases.

<i>Database</i>	<i>Ref.</i>	<i>Link</i>
NanoE-Tox	[24]	https://kbfi.ee/
ECOTOX	[37]	https://cfpub.epa.gov/ecotox/
NORMAN	[45]	https://www.norman-network.com/nds/ecotox/
DaNa	[14]	http://nanopartikel.info/
ERA	[23]	Not available.
EAT	[21]	http://www.ecetoc.org/
NIL	[16]	http://nanoparticlibrary.net/
NHECD	[46]	http://www.nhecd-fp7.eu/index.php?id=515
Nanomaterial Registry	[17]	http://www.nanomaterialregistry.org/
eNanoMapper	[32]	https://data.enanmapper.net/
nanoHub	[39]	https://nanohub.org/resources/databases
OCHEM	[41, 42]	https://ochem.eu/home/show.do
NECID	[30]	https://www.necid.eu/
NanoDataBank	[19]	http://nanoinfo.org/nanodatabank/
NanoMiner	[18]	https://ruho.uta.fi/wp/estools/nanommune/index.php/
S2NANO	[47]	http://portal.s2nano.org/dataRepository/database.do
NIOSH	[22]	https://www.cdc.gov/niosh/chemicals/dbsandtools.html
MOD-ENP-TOX	[23, 43]	https://fys.kuleuven.be/apps/modenptox/
CaNanoLab	[33]	https://cananolab.nci.nih.gov/caNanoLab/#/
ModNanoTox	[48]	https://www.birmingham.ac.uk/generic/modnanotox/index.aspx
NBI Knowledge Base	[40]	http://nbi.oregonstate.edu/library.php
NanoMILE	[28]	https://ssl.biomax.de/nanomile/cgi/login_bioxm_portal.cgi
NanoExPert	[44]	https://nanoeexpert.usace.army.mil/#/Pages/ToolSelectionPage.xaml
NanoWerk	[38]	http://www.nanowerk.com/nanomaterial-database.php
InterNano	[35]	https://www.internano.org/user/login
NTP Database	[20]	https://tools.niehs.nih.gov/ntp_tox/
TOXNET	[49]	https://www.nlm.nih.gov/toxnet/index.html
GUIDEnano	[31]	https://tool.guidenano.eu/
NIKC	[50]	http://nikc.egr.duke.edu/#!/dataset
Nanodatabase	[51]	http://nanodb.dk/en/search-database
NanoToxdb	[34]	http://iittrindia.org/envis/Default.aspx
Nano EHS	[25]	http://icon.rice.edu/report.cfm
NanoPUZZLES	[26]	https://zenodo.org/record/35493#.Xv5g0EBuK_s

included in the database. Raw data collected from analytical instruments are manually refined, validated, and organized before being stored in the MESOCOSM database.

The free access to the content of the MESOCOSM database and its integration into important databases such as eNanoMapper, NanoInformaTIX, ... is expected to provide a vast collection of environmental nanosafety data to support governments and regulators in developing legislation and recommendations regarding the use of ENMs. Furthermore, in a safer-by-design perspective, using the MESOCOSM database, manufacturers could better understand and assess the environmental exposure and toxic potential of ENMs at each stage of their life cycle.

3.2. MESOCOSM database content

The MESOCOSM database contains all the information collected during the contamination of an ecosystem by ENMs [3]. The simulation in the lab of this ecosystem is ensured by mesocosm testing, that offer a means of providing meaningful environmental exposure and hazard data to inform about environmental risk [11]. As described in Figure 1, this meaningful data can be divided into six groups:

- *The mesocosm testing information:* this group of information includes a general description of the conditions and the design of the experiment, such as the Digital Object Identifier (DOI) of the publication in which the results of the experiment were published, the total duration of the experiment, the type of ecosystem to be studied, the primary and secondary consumers, the benthic and planctonic primary producers representing the food chain of the mimicked ecosystem.
- *The other additional information:* This group of information represents more general information on the peer-reviewed publications associated with the data, such as details about the publication in which the results of the experiment were published, details about the authors of the publication, usually those who conducted the experiment, and details about the institutions to which the authors belong.
- *The physicochemical properties:* This set of parameters provides details of the physical and chemical properties of the nanoparticle used as a contaminant. This group of parameters can include the chemical nature, mineralogy, shape, coating, and initial size of the nanoparticle. This group also includes the name of the contaminant which is specified by concatenating its physicochemical properties, e.g. "Ag_Sphere_Citrate_20" corresponds to a silver (Ag), spherical, citrate-coated ENM with a size of 20 nm. Another item included is the sampling time representing the time for collecting the values from samples to determine the experimental parameters [10].
- *The environmental end-points:* This set of parameters includes indicators for assessing the state of the environment as function of time, such as the pH, temperature, conductivity, dissolved oxygen, total organic carbon (TOC), oxidation-reduction potential (ORP) in the water column, surficial sediment, and at the sediment-water interface, and the number of suspended particles in the water column.
- *The exposure end-points:* this set includes specific parameters used to measure and estimate the exposure of the studied environment to metal-based ENMs as function of time, such as the total concentration of metal in the surficial sediment, and water column, dissolved concentration of metal in the water column, ENMs size in the column water, total concentration of metal in layings, juveniles, adults, and digestive glands of adults, as well as the speciation in the water, sediment, layings, juveniles, adults, and digestive gland.
- *The hazard end-points:* This group of parameters includes biomarkers to measure observable changes indicating the hazard (acute or chronic) of ENMs towards living organism, such as the number of adult and juveniles organisms, number of picoplankton and picobenthos, number of algae in the water column and surficial sediment, thiobarbituric acid reactive substances (TBARS) or the total antioxidant capacity (TAOC).

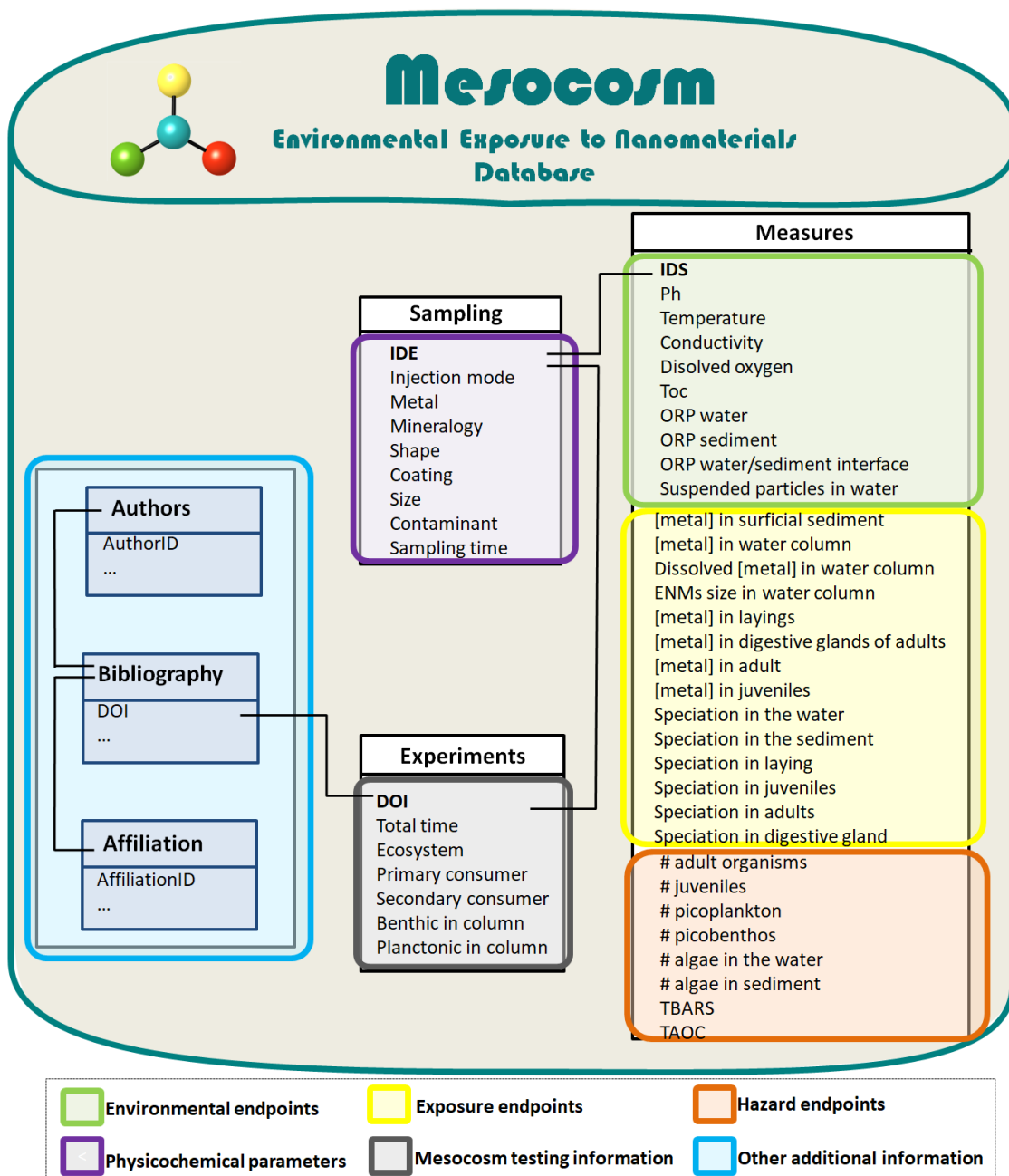


Figure 1: The MESOCOSM database management system content.

3.3. MESOCOSM database development

The MESOCOSM database system is implemented using JAVA programming language with MySQL. It has been developed into two levels, the back-end and the front-end layers. The back-end layer consists of the database server implementation, and the front-end layer consists of the graphical user interface. The back-end layer was firstly built on an Apache HTTP Server 2.4.4.1 with MySQL server. It was then hosted on an external server to be accessible to all users. The relational database was developed using the Structured Query Language (SQL) through the relational database management MySQL version 8.0. The interface

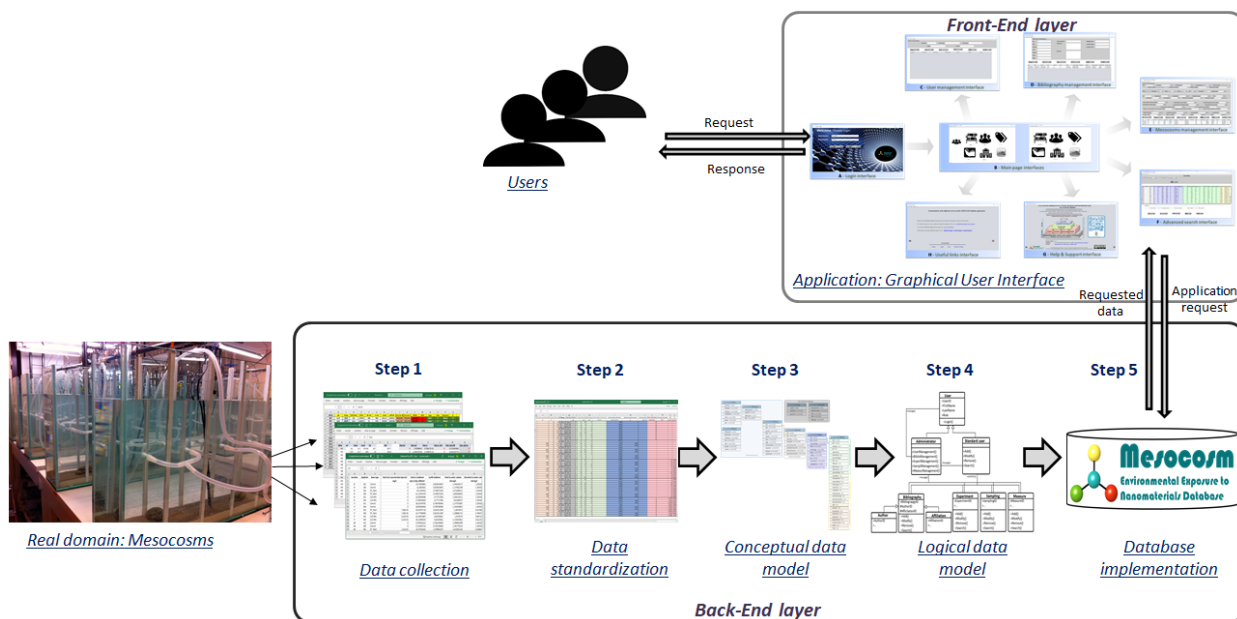


Figure 2: Methodology followed for designing and implementation of the MESOCOSM database management system.

component (front-end layer) is designed and implemented in JAVA programming language (Java SE 13.0.1). It allows users to access and manipulate data and services hosted by MESOCOSM database server. In addition, a web page has been developed especially for the MESOCOSM database system. This website provides a thorough description of MESOCOSM, its data, structure, some help documents, and links to download its user interface and its source files. This web page is implemented in HTML, CSS, JavaScript, and PHP languages, and is hosted on the GitHub repository.

The methodological steps followed during the implementation of the MESOCOSM database system are presented in Figure 2. In the following sections, each part of the methodology is described in detail.

• Step 1: Data collection

This first phase consists of collecting all the quantitative and qualitative data regarding the spatial and temporal behavior, fate and impact of ENMs obtained during mesocosm experiments. Raw data collected from analytical instruments are manually refined, validated, and organized. These data describe the exposure, the hazard, and the bio-physico-chemical properties of the ecosystems mimicked. All the files collected in this step contain data that are heterogeneous, not synchronized and collected from different individuals each one having its own style of structuring its information and its own domain of expertise.

• Step 2: Data integrity and standardization

Collected files were pre-processed to ensure their reliability, availability, maintainability, and efficiency. The experts propose a consistent data structure that is as general as possible without losing or affecting the quality of the data. This standard file prescribes the format that the team members must absolutely respect, and by which their collected information can be imported and stored in the MESOCOSM database. Accordingly, raw data collected from analytical instruments are manually processed and organized in this file before being stored in the MESOCOSM database. The standard file mainly focuses on the ENMs properties, environmental exposure and hazard endpoints.

• Step 3: The conceptual data model

As illustrated by Figure 3, the conceptual data model represents the schema of the database, i.e. the most important entities and their relationships. An entity is a real-world object, while a relationship is an association or a dependency between two entities. Based on entity-relationship diagrams, the objective of the conceptual model is to communicate contextual knowledge to any individuals who are unfamiliar to the domain. The conceptual data model presents also some detail about the entities such as their attributes (the properties that characterize an entity set), their cardinality (the relationships between tables), and their constraints (the conditions forced on the columns of the table to meet the data integrity).

• Step 4: The logical data model

As opposed to a conceptual data model, which may have very general terms, the logical data model includes a further level of detail, supporting both the architecture of the application and data requirements. It defines the primary keys, foreign keys, alternate keys, and inversion entities of the database. Table 2 presents the optimized logical data model of the MESOCOSM database which consists of nine tables. This logical data model uses indexes (primary keys) and foreign keys to represent data relationships, but these are defined in a generic database context independent of any specific database management system tool. For example, the user number (*userID*) represents the index of the *USERS* table and can uniquely identify a user. To distinguish foreign keys, a prefix consists of the sharp sign (#) is added to the beginning of the attribute(s) that represent(s) them. For example, the Digital Object Identifier (*DOI*) is the primary key of the *PUBLICATION* table and is also used as a foreign key in the *BIBLIOGRAPHY* table #*DOI* attribute to ensure the relationship between these two tables.

Table 2: The logical modeling of the MESOCOSM database management system.

Table	Description
1	Users (<i>userID</i> , <i>firstName</i> , <i>lastName</i> , <i>role</i> , <i>assignment</i>).
2	Bibliography (<i>bibliographyID</i> , # <i>authorID</i> , # <i>affiliationID</i> , # <i>DOI</i>).
3	Publication (<i>DOI</i> , <i>title</i> , <i>journal</i> , <i>year</i> , <i>volume</i> , <i>issue</i> , <i>page</i> , <i>contactAuthor</i> , <i>keywords</i> , <i>abstract</i> , <i>publishStatus</i>).
4	Author (<i>authorID</i> , <i>firstName</i> , <i>lastName</i> , <i>middleName</i> , <i>email</i> , <i>phone</i> , <i>web</i>).
5	Description (<i>descriptionID</i> , <i>tableDescription</i> , <i>field</i> , <i>note</i> , <i>dataType</i>).
6	Affiliation (<i>affiliationID</i> , <i>department</i> , <i>institution</i> <i>city</i> , <i>state</i> , <i>zipcode</i> , <i>country</i>).
7	Experiment (<i>experimentID</i> , # <i>DOI</i> , <i>totalDose</i> , <i>totalDuration</i> , <i>ecosystemType</i> , <i>injectionMode</i> , <i>primaryConsumer</i> , <i>secondaryConsumer</i> , <i>benthicInoculum</i> , <i>planktonicInoculum</i>).
8	Sampling (<i>samplingID</i> , # <i>experiment</i> , <i>metal</i> , <i>mineralogy</i> , <i>shape</i> , <i>coating</i> , <i>size</i> , <i>contaminant</i> , <i>samplingTime</i>).
9	Measure (<i>measureID</i> , # <i>sampling</i> , # <i>experiment</i> , <i>pH</i> , <i>temperature</i> , <i>conductivity</i> , <i>dissolvedOxygen</i> , <i>TOC</i> , <i>ORPwater</i> , <i>ORPsediment</i> , <i>ORPWaterSedimentInterface</i> , <i>numberOfParticlesInWater</i> , <i>ENMsSizeInWater</i> , <i>totalConcentrationOfMetalInSurficialSediment</i> , <i>totalConcentrationOfMetalInWaterColumn</i> , <i>dissolvedConcentrationOfMetalInWaterColumn</i> , <i>totalConcentrationOfMetalInWholeLayings</i> , <i>totalConcentrationOfMetalInTheDigestiveGland</i> , <i>totalConcentrationOfMetalInWholeAdult</i> , <i>totalConcentrationOfMetalInWholeJuveniles</i> , <i>changeInMetalSpeciationInTheDigestiveGland</i> , <i>changeInMetalSpeciationInLayings</i> , <i>changeInMetalSpeciationInJuveniles</i> , <i>numberOfAdultOrganisms</i> , <i>numberOfJuvenilesOrganisms</i> , <i>numberOfPicoplankton</i> , <i>numberOfPicobenthos</i> , <i>numberOfAlgaeInWater</i> , <i>numberOfAlgaeInSediment</i> , <i>TBARS</i> , <i>TAOC</i>).

• Step 5: The database implementation

The purpose of this step is to convert the logical model to a physical prototype. It consists in developing the programs to create the database and populate it. Once the database is implemented, numerous tests are performed to detect bugs or programming errors. This phase is expected to produce all the files that will be made available to the final user, including the programs, user manual and configuration files.

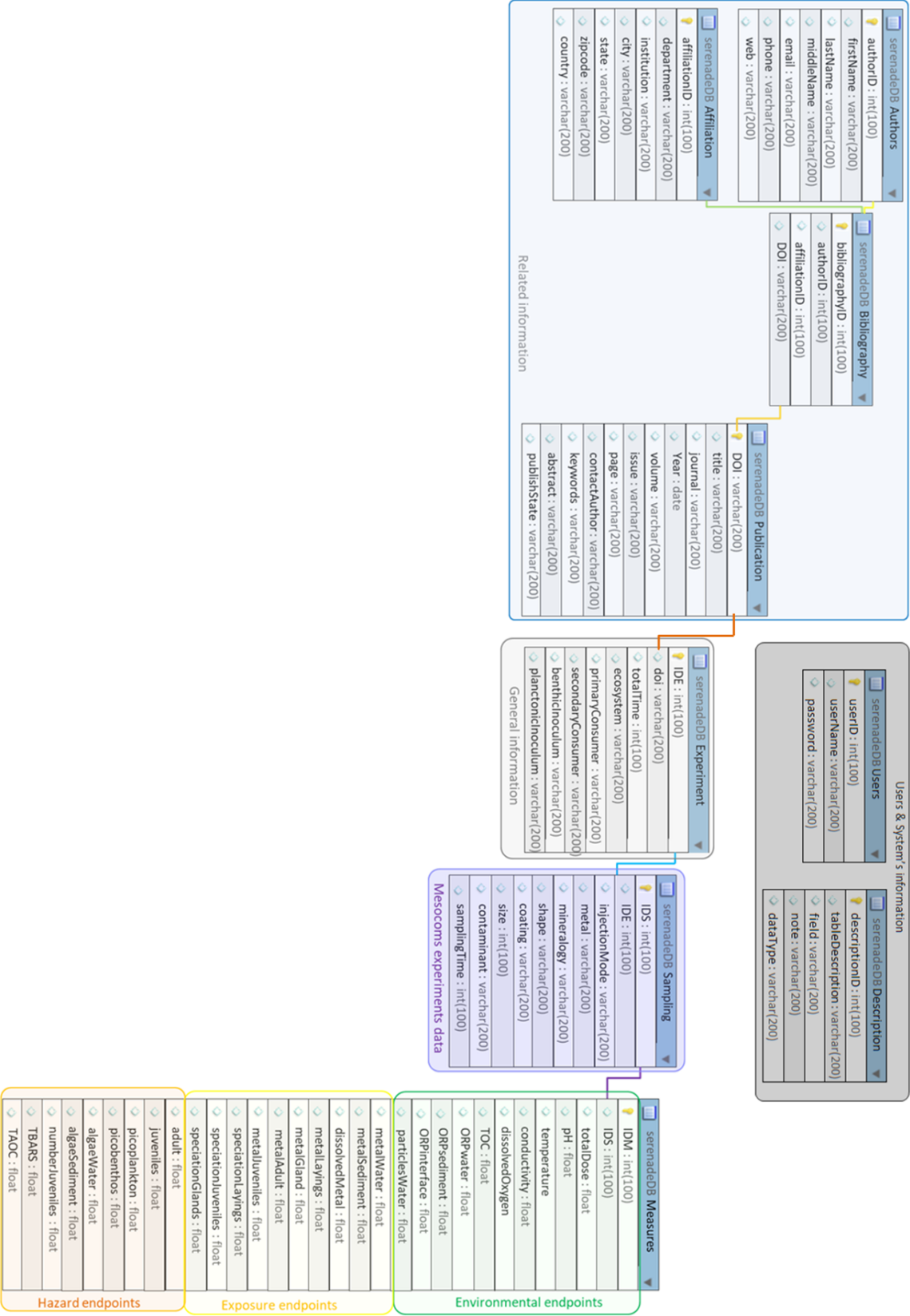


Figure 3: The conceptual data model of the MESOCOSM database system.

3.4. MESOCOSM database interface

The MESOCOSM user interface allows users to access and add new data in the MESOCOSM database system. This proposed interface provides a logical interface to the MESOCOSM database management system functionalities and specifies how users should interact with it. It has been developed as an open source JAVA library. Figure 4 presents the UML class diagram describing the architecture of our user interface by modeling its classes, attributes, methods, and relationships among the classes involved in our application. A class is represented by a rectangle separated into three parts (e.g. the *Users* class): (i) the first part contains the name of the class (e.g. "*Users*"), (ii) the second part contains the attributes of the class (e.g. "*userID*", "*firstName*", etc.), and (iii) the last part contains the methods of the class (e.g. *Log-in()*).

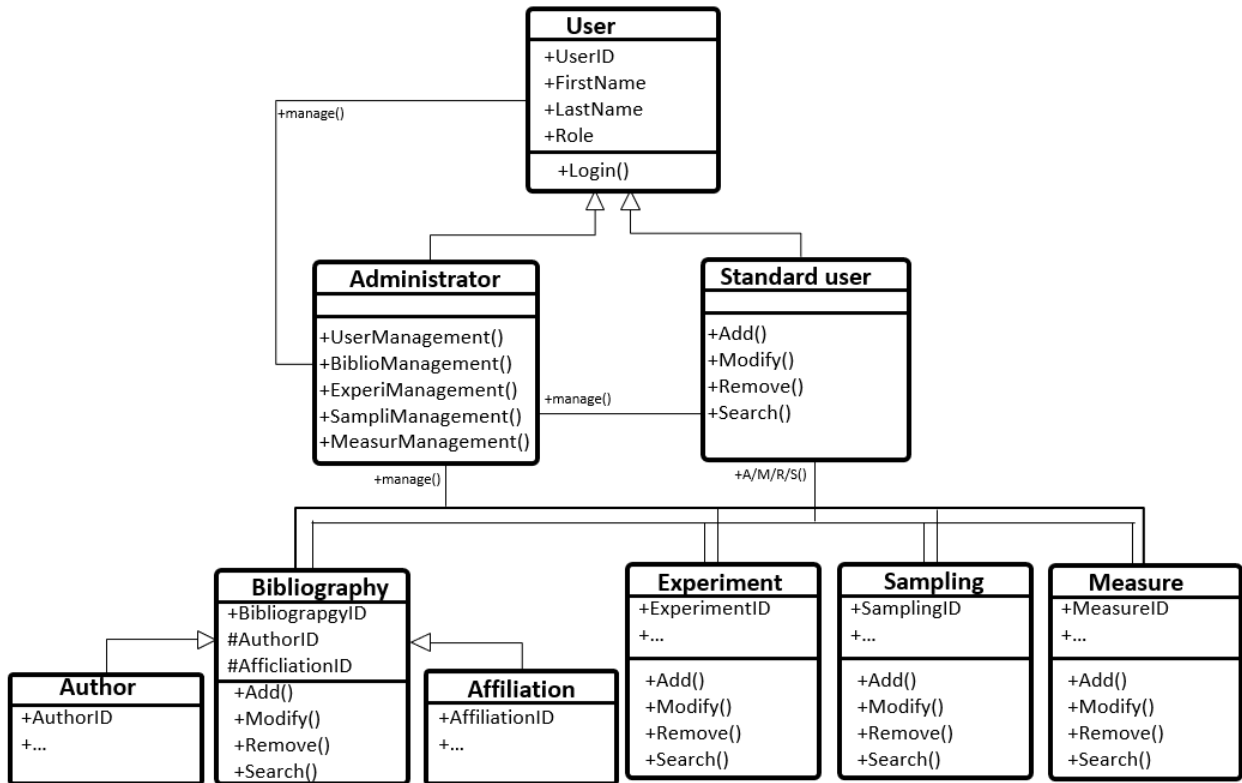


Figure 4: The class diagram representing the architecture of the proposed MESOCOSM database management system.

The main classes of our application are:

- *Users* class contains all the users of the application.
- *Standard user* class includes all the activities handled by the internal user, such as Adding / Modifying / Removing / Searching bibliographies, publications, experiments, sampling, and measures.
- *Administrator* class includes all the methods supported by the administrator, such as manage users, bibliographies, publications, affiliations, authors, experiments, sampling, and measures.
- *Bibliography* class contains information on the peer-reviewed publications associated with the data (i.e. DOI, authors, and affiliation).

- *Sampling* class includes precise measurements on the experimental conditions of the mesocosm experiments (the sampling time, total duration of the experiment, type of ecosystem to be studied), and the physicochemical properties related to their contaminants (their size, surface, ...).
- *Measures* class contains the measurements related to the experiments such as the environmental, exposure and hazard endpoints.

The different classes forming our application are organized in two main packages, the *PackInterface* and *PackBase* packages. The *PackInterface* package contains all the classes dedicated to the *JavaSwing* classes, and a class gathering the static methods of interaction among these classes (the display calls). The *PackBase* package includes the classes that manage the database connection, model the queries, define the objects of the model that can be modified by the application.

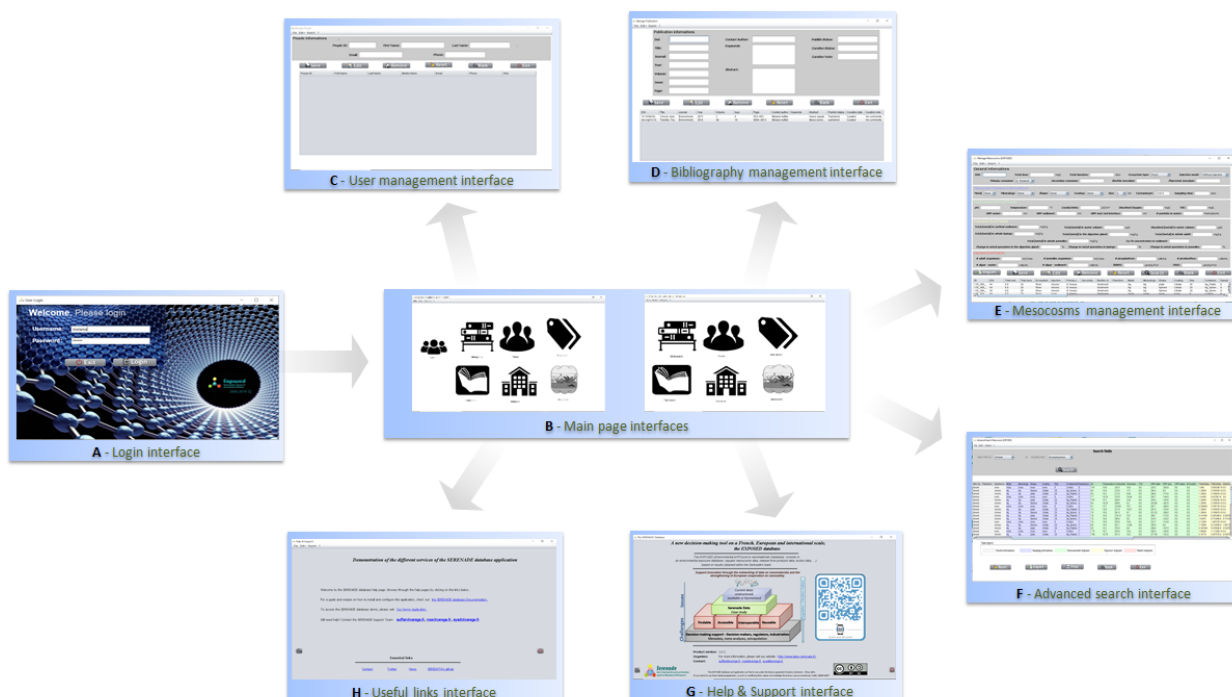


Figure 5: User interface of the MESOCOSM database management system. (A) The login page. (B) The home page: a specific home page for each user according to the user role. (C) The user account management page. (D) The bibliography management page. (E) The mesocosms management page. (F) The advanced search page. (G) The help & support page, and (H) The useful links page.

The MESOCOSM database management system is accessible via a user interface presented in Figure 5. Using this interface users can have access to the whole or part of the dataset depending on their needs. For instance, in their recent paper Nassar et al. [52] extracted from the database exposure and hazard data obtained in indoor aquatic mesocosms mimicking a pond ecosystem contaminated to CeO_2 ENMs. Applying multivariate analysis on these extracted data, they concluded that the exposure scenario constituted the main parameter to consider while investigating the risk for a given ecosystem. Following a single pulse dosing of CeO_2 ENMs, the global response of the pond ecosystem was time dependent, whereas a multiple dosing contamination failed to significantly perturb the system over a month. Interestingly, they highlighted that the ENMs surface coating played a secondary role and affect the global response of the pond ecosystem only in the short term only.

4. How fair is MESOCOSM ?

MESOCOSM was designed from the outset with maximum adherence to the FAIR principles that promote finding, accessing, interoperating, and reusing shared data [53]. From the beginning, the MESOCOSM database system for environmental nanosafety was designed following the FAIR data vision to optimize environmental exposure data sharing and reuse by humans and machines. The proposed database system considers also the principles of linked data mentioned in the works of Bizer et al. [54, 55] demonstrating that each entity must be correctly identified with a unique identifier, described in a semantic format, and linked with other resources. The idea from the beginning was to identify all the records with a persistent and unique identifier (PID) so that they can be located unambiguously by machines, and easily findable by humans through the richness of the metadata used to describe them. Moreover, the database content is registered and archived in different websites (the CEREGE website, GitHub repository¹, and Perma web archiving service²) so that it can be easily found and accessible. Regardless of what may happen to the original source, the archived record will always be available through the cited websites. Similarly, all datasets have been published and are identified with unique DOI identifiers. To make the MESOCOSM database content interoperable and FAIRness for machines, we have adopted a semantic format, the Resource Description Framework (RDF). This semantic format is used directly as the primary medium for linked data in MESOCOSM. Indeed, all the entities in the database are linked to other resources with permanent unique URI identifiers through an ontology especially designed to support the development of the MESOCOSM database system for environmental nanosafety. These identifiers can be considered unique across the World Wide Web and therefore provide a solid basis for linking the MESOCOSM datasets. Indeed, each data or entry in the database is associated with an ontological concept that defines its meaning. Indeed, each data or entry in the database is associated with an ontological concept that defines its meaning so that the datasets can be sufficiently understood by humans and can be reused appropriately. In addition, the RDF format makes the MESOCOSM data accessible in the semantic web through the powerful SPARQL query mechanism and rich metadata indexing. The reusability of the MESOCOSM database system is also ensured by the use of an open source license. The default license of MESOCOSM is Creative Commons (CC-BY 4) which allows users to download, modify and reuse the data without restriction, but attribution of the source must accompany the reuse. The adoption of the four FAIR characteristics, makes MESOCOSM a modern database making environmental exposure data valuable to a wide range of users, including manufacturers, researchers, and government agencies.

5. Conclusion and perspectives

In this paper, we present MESOCOSM, a mesocosm database management system for environmental nanosafety, developed to provide a collection of exposure and impact data to ENMs experimentally measured using aquatic mesocosms. The database can be used in many different applications aiming to assess and predict the ENMs behavior and fate in different ecosystems as well as their potential impacts on the environment at different stages of the nanoproductions lifecycle. The data provided by MESOCOSM will actively participate in optimizing the risk/benefit ratio over the entire life cycle of ENMs, from the manufacture of an industrial product to its end of life, and, therefore support the "safer by design" approach which aims to minimize the environmental exposure or hazards of ENMs.

Currently, all the entries stored in the MESOCOSM are published data obtained in mesocosms by the SERENADE consortium. The MESOCOSM database will be frequently populated with new data obtained in mesocosms facilities provided by groups working on nanosafety (as the CEINT center) but also, with other literature data. Indeed, we are focusing on the exploration and use of text mining techniques for enriching the MESOCOSM database. As such, we already proposed an ontology-based Natural Language Processing (NLP) approach to automatically extract and transfer data from text sources (typically scientific

¹It offers provides hosting for software development and version control using Git command-line interface

²Perma.cc is a web archiving service for legal and academic citations created and maintained by the Harvard Law School Library

articles and expert reports) [56]. The proposed system, firstly, exploits the NLP techniques to intelligently analyze textual documents and identify pertinent information from document [57], and secondly, uses a domain ontology to semantically classify and categorize the extracted information according to its concepts, thus associating the information with their corresponding attributes within the database. We demonstrated how the automatic extraction system can extract environmental exposure to ENMs information from text sources, classify them according to the ontology concepts, and automatically enrich the database. This will significantly strengthen the MESOCOSM database by data coming from others groups using other mesocosm designs.

Short-term perspectives would be the incorporation of additional user-friendly tools such as data analytics tools based on some innovative algorithms for predicting the potential impact of ENMs on the environmental on the long term or for categorizing the ENMs based on the potential risk for the environment.

Based on the relational model of the MESOCOSM database, an ontology is currently under development. This ontology will provide the definition of vocabulary and specifies the meaning (semantics) of the MESOCOSM database entries. The development of such ontology will improve the interconnectedness and interoperability of the model, and therefore facilitate both the access and the interrogation of SERENADE and CEINT datasets. In addition, the ontology will contribute to the extension of one of the most important European nanosafety ontology, the eNanoMapper [32], to cover another major dimension of nanosafety: Environmental Exposure.

Acknowledgments

This work is a contribution to the Labex Serenade (No. ANR-11-LABX-0064) funded by the “Investissements d’Avenir” French Government program of the French National Research Agency (ANR) through the A MIDEX project (No. ANR-11-IDEX-0001-02), and a contribution to the European project NanoInformaTIX, H2020-NMBPTO-IND-2018-2020-814426. This work is also a contribution to the OSU-Institut Pythéas. The authors acknowledge the CNRS for the funding of the IRP iNOVE, and the Sustainable Environment Group of the CEREGE for helpful discussions and valuable feedback.

Conflicts of interest

The authors report no conflicts of interest.

References

- [1] A. Zabeo, J. M. Keisler, D. Hristozov, A. Marcomini, I. Linkov, Value of information analysis for assessing risks and benefits of nanotechnology innovation, *Environmental Sciences Europe* 31 (1) (2019) 11.
- [2] P. Kangas, W. Adey, Mesocosm management.
- [3] W. Adey, P. Kangas, Greenhouses, microcosms, and mesocosms, *Ecosystem Ecology* (2008) 281.
- [4] K. D. Kimball, S. A. Levin, Limitations of laboratory bioassays: the need for ecosystem-level testing, *Bioscience* 35 (3) (1985) 165–171.
- [5] S. Karcher, E. L. Willighagen, J. Rumble, F. Ehrhart, C. T. Evelo, M. Fritts, S. Gaheen, S. L. Harper, M. D. Hoover, N. Jeliaskova, et al., Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations, *NanoImpact* 9 (2018) 85–101.
- [6] T. Marie, A. Melanie, B. Lenka, I. Julien, K. Isabelle, P. Christine, M. Elise, S. Catherine, A. Bernard, A. Ester, et al., Transfer, transformation, and impacts of ceria nanomaterials in aquatic mesocosms simulating a pond ecosystem, *Environmental science & technology* 48 (16) (2014) 9004–9013.
- [7] M. Auffan, M. Tella, C. Santaella, L. Brousset, C. Paillès, M. Barakat, B. Espinasse, E. Artells, J. Issartel, A. Masion, et al., An adaptable mesocosm platform for performing integrated assessments of nanomaterial risk in complex environmental systems, *Scientific reports* 4 (2014) 5608.
- [8] A. Kraegeloh, B. Suarez-Merino, T. Sluijters, C. Micheletti, Implementation of safe-by-design for nanomaterial development and safe innovation: why we need a comprehensive approach, *Nanomaterials* 8 (4) (2018) 239.
- [9] E. Valsami-Jones, I. Lynch, C. Charitidis, Nanomaterial ontologies for nanosafety: a rose by any other name..., *J. Nanomed. Res.* 3 (2016) 00070.
- [10] J. Bottero, J. Rose, C. De Garidel, A. Masion, T. Deutsch, G. Brochard, M. Carrière, N. Gontard, H. Wortham, T. Rabiloud, et al., Serenade: safer and ecodesign research and education applied to nanomaterial development, the new generation of materials safer by design, *Environmental Science: Nano* 4 (3) (2017) 526–538.

- [11] M. Auffan, A. Masion, C. Mouneyrac, C. de Garidel-Thoron, C. O. Hendren, A. Thiery, C. Santaella, L. Giamberini, J.-Y. Bottero, M. R. Wiesner, et al., Contribution of mesocosm testing to a single-step and exposure-driven environmental risk assessment of engineered nanomaterials, *NanoImpact* 13 (2019) 66–69.
- [12] M. Tella, M. Auffan, L. Brousset, E. Morel, O. Proux, C. Chanéac, B. Angeletti, C. Pailles, E. Artells, C. Santaella, et al., Chronic dosing of a simulated pond ecosystem in indoor aquatic mesocosms: fate and transport of ceo 2 nanoparticles, *Environmental Science: Nano* 2 (6) (2015) 653–663.
- [13] M. Auffan, W. Liu, L. Brousset, L. Scifo, A. Pariat, M. Sanles, P. Chaurand, B. Angeletti, A. Thiéry, A. Masion, et al., Environmental exposure of a simulated pond ecosystem to a cuo nanoparticle-based wood stain throughout its life cycle, *Environmental Science: Nano* 5 (11) (2018) 2579–2589.
- [14] H. F. Krug, N. Bohmer, D. Kühnel, C. Marquardt, K. Nau, C. Steinbach, The dana2. 0 knowledge base nanomaterials—an important measure accompanying nanomaterials development, *Nanomaterials* 8 (4) (2018) 204.
- [15] C. Marquardt, D. Kühnel, V. Richter, H. Krug, B. Mathes, C. Steinbach, K. Nau, Latest research results on the effects of nanomaterials on humans and the environment: Dana—knowledge base nanomaterials, in: *Journal of Physics: Conference Series*, Vol. 429, IOP Publishing, 2013, p. 012060.
- [16] A. L. Miller, M. D. Hoover, D. M. Mitchell, B. P. Stapleton, The nanoparticle information library (nil): a prototype for linking and sharing emerging data, *Journal of Occupational and Environmental Hygiene* 4 (12) (2007) D131–D134.
- [17] K. C. Mills, D. Murry, K. A. Guzan, M. L. Ostraat, Nanomaterial registry: database that captures the minimal information about nanomaterial physico-chemical characteristics, *Journal of nanoparticle research* 16 (2) (2014) 2219.
- [18] L. Kong, S. Tuomela, L. Hahne, H. Ahlfors, O. Yli-Harja, B. Fadeel, R. Lahesmaa, R. Autio, Nanominer—integrative human transcriptomics data resource for nanoparticle research, *PLoS One* 8 (7) (2013) e68414.
- [19] Nanodatabank.org, nanodatabank website, <http://nanoinfo.org/nanodatabank/> (cited July 2020).
- [20] Ntp.org, Ntp website, <https://ntp.niehs.nih.gov/data/index.html> (cited July 2020).
- [21] Ecetoc.org, Ecetoc website, <http://www.ecetoc.org/publication/tr-091-ecetoc-aquatic-toxicity-eat-database-eat-database/> (cited July 2020).
- [22] Niosh.org, niosh website, <https://www.cdc.gov/niosh/chemicals/dbsandtools.html> (cited July 2020).
- [23] H. Vriens, D. Mertens, R. Regret, P. Lin, J.-P. Locquet, P. Hoet, Case study iii: The construction of a nanotoxicity database—the mod-enp-tox experience, in: *Modelling the Toxicity of Nanoparticles*, Springer, 2017, pp. 325–344.
- [24] K. Juganson, A. Ivask, I. Blinova, M. Mortimer, A. Kahru, Nanoe-tox: New and in-depth database concerning ecotoxicity of nanomaterials, *Beilstein journal of nanotechnology* 6 (1) (2015) 1788–1804.
- [25] J. Youtie, A. Porter, P. Shapira, L. Tang, T. Benn, The use of environmental, health and safety research in nanotechnology research, *Journal of nanoscience and nanotechnology* 11 (1) (2011) 158–166.
- [26] A.-N. Richarz, A. Avramopoulos, E. Benfenati, A. Gajewicz, N. G. Bakhtyari, G. Leonis, R. L. M. Robinson, M. G. Papadopoulos, M. T. Cronin, T. Puzyn, Compilation of data and modelling of nanoparticle interactions and toxicity in the nanopuzzles project, in: *Modelling the Toxicity of Nanoparticles*, Springer, 2017, pp. 303–324.
- [27] R. L. M. Robinson, M. T. Cronin, A.-N. Richarz, R. Rallo, An isa-tab-nano based data collection framework to support data-driven modelling of nanotoxicology, *Beilstein journal of nanotechnology* 6 (1) (2015) 1978–1999.
- [28] A. Biomax Informatics, The nanomile knowledge base (2018).
- [29] E. Joossens, P. Macko, T. Palosaari, K. Gerloff, I. Ojea-Jiménez, D. Gilliland, J. Novak, S. F. Torrent, J.-M. Gineste, I. Römer, et al., A high throughput imaging database of toxicological effects of nanomaterials tested on hepg cells, *Scientific data* 6 (1) (2019) 1–10.
- [30] J. Pelzer, Structure and functionality of the nano exposure and contextual information database (necid), *GEFAHRSTOFFE REINHALTUNG DER LUFT* 73 (7-8) (2013) 302–304.
- [31] M. L. Fernández-Cruz, D. Hernandez-Moreno, J. Catalán, R. Cross, H. Stockmann-Juvala, J. Cabellos, V. R. Lopes, M. Matzke, N. Ferraz, J. J. Izquierdo, et al., Quality evaluation of human and environmental toxicity studies performed with nanomaterials—the guidenano approach, *Environmental Science: Nano* 5 (2) (2018) 381–397.
- [32] N. Jeliaskova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliaskov, N. Kochev, et al., The enanomap database for nanomaterial safety information, *Beilstein journal of nanotechnology* 6 (1) (2015) 1609–1634.
- [33] S. A. Morris, S. Gaheen, M. Lijowski, M. Heiskanen, J. Klemm, cananolab: A nanomaterial data repository for biomedical research, in: *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2014, pp. 29–33.
- [34] Nanotox.org, nanotoxdb website, <http://iittrindia.org/envis/Default.aspx> (cited July 2020).
- [35] J. Adamick, Internano: Serving the nanomanufacturing community.
- [36] R. Reznik-Zellen, B. Stevens, M. Thorn, J. Morse, M. Smucker, J. Allan, D. Mimno, A. McCallum, M. Tuominen, Internano: e-science for the nanomanufacturing community, in: *2008 IEEE Fourth International Conference on eScience*, IEEE, 2008, pp. 382–383.
- [37] C. Elonen, Ecotox: Ecotoxicology knowledgebase system user guide – version 5.1, EPA Office of Research and Development EPA/600/R-18/377.
- [38] nanowerk.com, nanowerk website, <http://www.nanowerk.com/nanomaterial-database.php> (cited July 2020).
- [39] Nanohub.org, nanohub website, <https://nanohub.org/resources/databases> (cited July 2020).
- [40] S. Harper, Nanomaterial-biological interactions knowledgebase.
- [41] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, et al., Online chemical modeling environment (ochem): web platform for data storage, model development and publishing of chemical information, *Journal of computer-aided molecular design* 25 (6) (2011) 533–554.
- [42] I. Oprisiu, S. Novotarskyi, I. V. Tetko, Modeling of non-additive mixture properties using the online chemical database and modeling environment (ochem), *Journal of cheminformatics* 5 (1) (2013) 1–7.

- [43] M. Jeroen Herczeg, Mod-enp-tox — home, <https://fys.kuleuven.be/apps/modenptox/> (2020).
- [44] NanoExPERTdb.org, Nanoexpert website, <https://nanoexpert.usace.army.mil/#/Pages/ToolSelectionPage.xaml> (cited July 2020).
- [45] W. Brack, V. Dulio, J. Slobodnik, The norman network and its activities on emerging environmental substances with a focus on effect-directed analysis of complex environmental contamination, *Environmental Sciences Europe* 24 (1) (2012) 29.
- [46] O. Maimon, A. Browarnik, Nhecd-nano health and environmental commented database, in: *Data mining and knowledge discovery handbook*, Springer, 2009, pp. 1221–1241.
- [47] S. project, S2nano website, <http://portal.s2nano.org/dataRepository/database.do> (2020).
- [48] Modnanoto.org, modnanotox website, <https://www.birmingham.ac.uk/generic/modnanotox/index.aspx1> (cited July 2020).
- [49] P. Wexler, Toxnet: an evolving web resource for toxicology and environmental health information, *Toxicology* 157 (1-2) (2001) 3–10.
- [50] C. O. Hendren, G. V. Lowry, J. M. Unrine, M. R. Wiesner, A functional assay-based strategy for nanomaterial risk forecasting, *Science of the Total Environment* 536 (2015) 1029–1037.
- [51] S. F. Hansen, K. A. Jensen, A. Baun, Nanoriskcat: a conceptual tool for categorization and communication of exposure potentials and hazards of nanomaterials in consumer products, *Journal of nanoparticle research* 16 (1) (2014) 2195.
- [52] M. Nassar, M. Auffan, C. Santaella, A. Masion, J. Rose, Multivariate analysis of the exposure and hazard of ceria nanomaterials in indoor aquatic mesocosms, *Environmental Science: Nano*.
- [53] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, *Scientific data* 3 (1) (2016) 1–9.
- [54] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: *Semantic services, interoperability and web applications: emerging concepts*, IGI Global, 2011, pp. 205–227.
- [55] C. Pommier, C. Michotey, G. Cornut, P. Roumet, E. Duchêne, R. Flores, A. Lebreton, M. Alaux, S. Durand, E. Kimmel, et al., Applying fair principles to plant phenotypic data management in gnpis, *Plant phenomics* 2019 (2019) 1671403.
- [56] A. Ayadi, M. Auffan, J. Rose, Ontology-based nlp information extraction to enrich nanomaterial environmental exposure database, *Procedia Computer Science* 176 (2020) 360–369.
- [57] A. Goker, J. Davies, The role of natural language processing in information retrieval: Searching for meaning and structure.