NanoPharos: Collection, curation & organization into machine-readable format of NMs data

D. Zouraris[1,2], A. Tsoumanis[1,3], A. G. Papadiamantis[1,2,4], G. Melgaraki[5], I. Lynch[2,4], A. Afantitis[1,2,3]

1. Introduction

NanoPharos (https://db.nanopharos.eu/) is a carefully designed database to store and organise data on the environmental health and safety of nanomaterials (nanoEHS). The datasets can be curated from literature, experimentally generated, or computationally derived either from 1st principles (so-called physics-based models) or using data-driven modelling approaches including machine learning, in accordance with the FAIR principles[1] (Fig. 1). The initial rationale of NanoPharos was to create a database to fill a gap for nanoinformatics modelling of nanomaterials properties, interactions, and impacts (adverse outcomes or mechanistic toxicity assessment utilising in vitro or in vivo assays). NanoPharos provides modelling-ready tab-delimited datasets that can be directly imported into computational workflows and software, such as the Isalos Analytics Platform or KNIME using, e.g., the Enalos+ KNIME nodes.

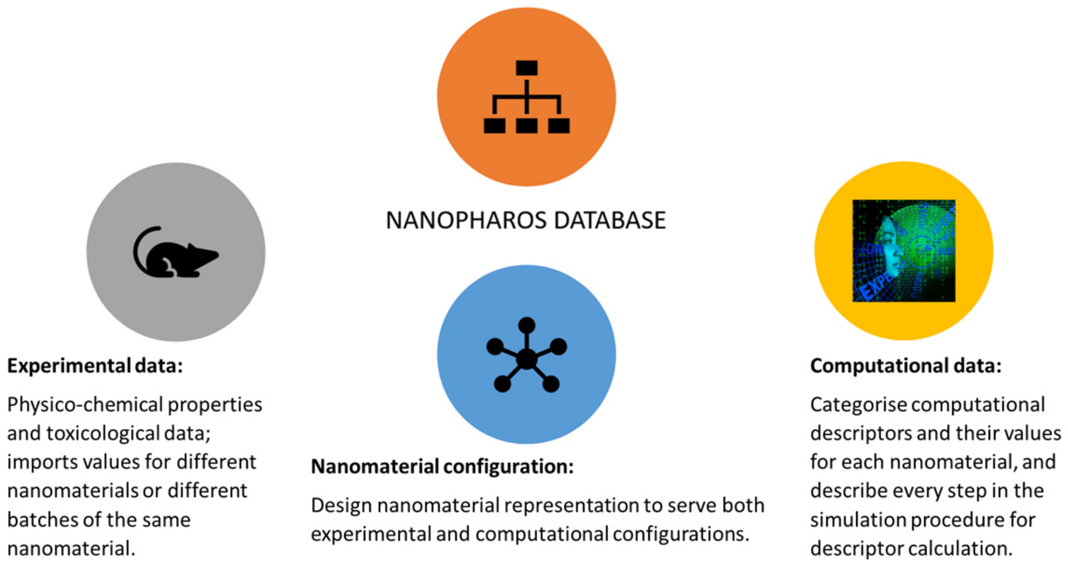

Figure 1: NanoPharos allows capture, storage, harmonisation and programmatic access to experimental and computational datasets relating to nanomaterials environmental health and safety (nanoEHS).

[1] NovaMechanics Ltd, Nicosia 1070, Cyprus; zouraris@novamechanics.com
[2] Entelos Institute, Larnaca 6059, Cyprus
[3] NovaMechanics MIKE, Piraeus 18545
[4] School of Geography, Earth and Environmental Sciences, University of Birmingham, B15 2TT, Birmingham, UK
[5] Division of Physical Sciences and Applications, Hellenic Military Academy, Vari 16672, Greece

2. NanoPharos Features

The NanoPharos database has several features and innovations that distinguish it from other databases, including:

Structured and FAIR-Compliant Framework

NanoPharos streamlines nanomaterials research with a FAIR-compliant framework, accelerating nanoEHS discoveries. Its high-quality ready for modelling datasets bridge a crucial nanoinformatics gap, enabling efficient and accurate analysis through platforms like KNIME (Fig. 2).

Dynamic Nanomaterials Management

NanoPharos adeptly tracks the variability and transformations of nanomaterials by defining separate batches and environmentally altered versions, linking them to a "parent" material. This tracking is supported by the European Materials Registry (ERM)[2] and NanoInChI identifiers which captures key features like composition and morphology[3].

Computational Analyses and Descriptor Enrichment

NanoPharos enhances nanomaterial data with computational descriptors for detailed characterization and facilitates *in silico* cytotoxicity prediction models. For instance, it used a dataset of 14 metal oxide nanomaterials with 62 descriptors to create a validated cytotoxicity model[4].

Inclusion of Omics Data and Comprehensive Data Spectrum

The extension of NanoPharos to include omics data marks a significant advance, covering the full spectrum of data needed for *in silico* exploitation of nanoEHS data. This comprehensive approach enables users to query and group datasets based on various criteria, facilitating detailed studies on nanomaterial impacts.

Iterative Improvement and Curator Engagement

NanoPharos evolves via continuous updates and curator collaboration, incorporating feedback and the latest research methods. An automated system for adding computational results into the database further enriches it with new insights and predictions.

3. Updates and Expansion Plans

NanoPharos plans to enhance its FAIR compliance and usability by introducing application programming interfaces (APIs) and KNIME nodes for better machine findability and actionability. Integrating ontologies like eNanoMapper and CheBI will make data and metadata machine understandable, fostering links between datasets and their corresponding nanoinformatics models, thus enhancing their FAIRness. Efforts are also being made to create a searchable registry for data and metadata, with the aim of certifying NanoPharos as a Trusted Repository under CoreTrustSeal or ISO16363 standards.

4. Recently Added Datasets and their added value to NanoPharos.

The addition of new datasets, including detailed examinations of Log*P* measurements for noble metal nanoparticles (NPs), PFOS adsorption efficiency by Gold NPs under varying pH conditions, reactive oxygen stress (ROS) induction by Gold NPs in HEK293 cells, zeta potential measurements for various NPs, and studies on the cellular uptake of NPs by A549 and HEK293 cells, represents a significant enrichment of the NanoPharos database.

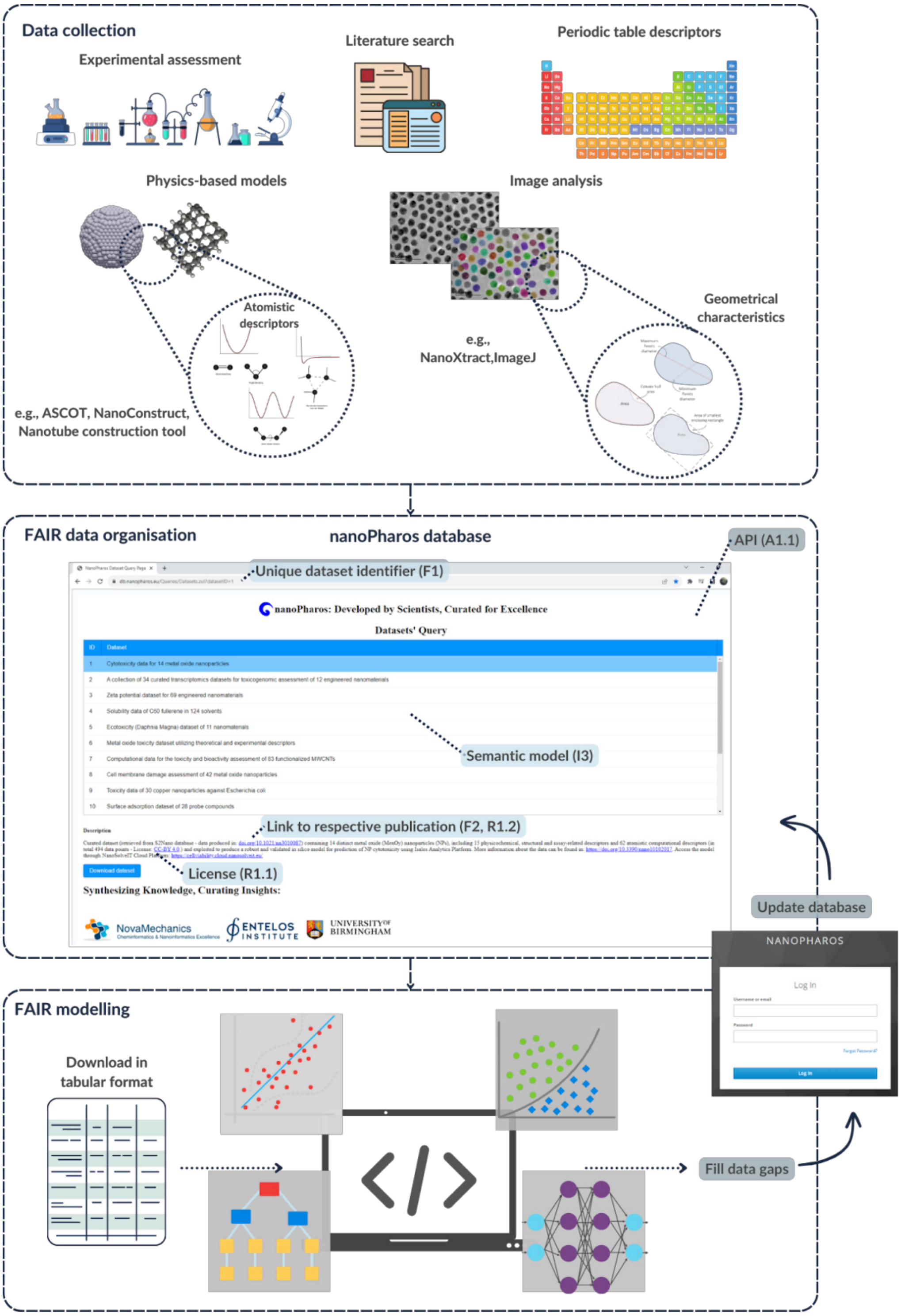

Figure 2: Schematic workflow for FAIRification of experimental and computational nanoEHS data via NanoPharos.

These datasets provide comprehensive physicochemical properties, bioactivity measures, and detailed insights into the interactions between nanomaterials and biological systems, serving several key purposes:

- **Enhanced Predictive Modelling Capabilities**: The computationally enrich datasets, enhance the accuracy of models predicting nanomaterial behaviour in environmental and biological contexts, advancing NanoPharos's support for research across environmental science to nanomedicine.
- **Facilitation of Interdisciplinary Research:** The structured, FAIR-compliant data format ensures their versatility and reusability across computational and analytical studies, reinforcing NanoPharos as a key nanomaterials research tool and promoting the safe development and use of nanotechnologies.
- **Promotion of Data Reusability:** NanoPharos offers open-access, harmonized data ready for modelling, greatly reducing entry barriers for researchers in nanomaterials safety and risk assessment, where gathering experimental data is often expensive and time-intensive.
- **Support for Environmental and Health Safety (EHS) Research**: Incorporating data on PFOS adsorption efficiency and ROS induction by nanomaterials provides essential insights into their environmental uses and potential toxic effects. This information is vital for creating nano-enabled solutions for water treatment, pollution reduction, and understanding of nanomaterials toxicity.
- **Advancement of Nanomedicine**: Datasets detailing the cellular uptake of nanomaterial by A549 and HEK293 cells are crucial for nanomedicine, shedding light on how physicochemical traits affect cellular interactions, essential for crafting efficient drug delivery systems, therapeutics, or diagnostics.

5. Conclusions

NanoPharos is an increasingly important database for nanoinformatics and nanosafety research, adhering to the FAIR (Findable, Accessible, Interoperable and Re-usable) principles for easy data access and sharing. It offers comprehensive tracking, detailed analysis, and diverse data, utilizing unique identifiers like ERM and NanoInChI for accuracy. Recent updates have enhanced its utility across various scientific fields, facilitating collaborative research, data reuse, and safety assessments. As it evolves, NanoPharos increasingly becomes vital for nanotechnology safety efforts.

This research was funded by the European Union's H2020 Marie Skłodowska-Curie Actions via CompSafeNano under (Grant Agreement n° 101008099).

6. References

(1) Wilkinson, M. D.; *et al.* Comment: The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* **2016**, *3*. shttps://doi.org/10.1038/sdata.2016.18.

(2) van Rijn, J.; *et al.* European Registry of Materials: Global, Unique Identifiers for (Undisclosed) Nanomaterials. *J Cheminform* **2022**, *14* (1). https://doi.org/10.1186/s13321-022-00614-7.

(3) Lynch, I.; *et al.* Can an InChI for Nano Address the Need for a Simplified Representation of Complex Nanomaterials across Experimental and Nanoinformatics Studies? *Nanomaterials* **2020**, *10*, 2493. https://doi.org/10.3390/nano10122493

(4) Papadiamantis, A. G.; *et al.* Predicting Cytotoxicity of Metal Oxide Nanoparticles Using Isalos Analytics Platform. *Nanomaterials* **2020**, *10* (10), 1–19. https://doi.org/10.3390/nano10102017.