NanoImpact 41 (2026) 100602

Contents lists available at ScienceDirect

NanoImpact

journal homepage: www.elsevier.com/locate/nanoimpact

Research Paper

nanoPharos: A case study on FAIR (Nano)material (Meta)data management

Anastasios G. Papadiamantis [a,b,*], Andreas Tsoumanis [b,c], Georgia Melagraki [d], Iseult Lynch [b,e,*], Antreas Afantitis [a,b,c,f,*]

[a] NovaMechanics Ltd., Nicosia, Cyprus
[b] Entelos Institute Ltd., Nicosia, Cyprus
[c] NovaMechanics MIKE, Piraeus, Greece
[d] Division of Physical Sciences and Applications, Hellenic Military Academy, Vari, Greece
[e] School of Geography, Earth, and Environmental Sciences, University of Birmingham, Birmingham, UK
[f] Department of Pharmacy, Frederick University, Nicosia, Cyprus

ARTICLE INFO

Editor: Bernd Nowack

Keywords:
Data repository
Database
FAIR (Meta)data management
Rich metadata
Advanced (Nano)materials
Ready-for-modelling datasets

ABSTRACT

Novel and advanced materials, including nanomaterials (NMs), are vital for diverse industrial and societal applications, yet conventional Research and Innovation (R&I) and Research and Development (R&D) can take decades to reach market deployment. Digitising these processes to support safe and sustainable material development, and reduce reliance on animal testing, requires large volumes of high-quality, interoperable data. The FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles provide a framework for this, but demand domain-specific implementation strategies. We present nanoPharos, a repository offering ready-for-modelling NMs datasets integrating physicochemical characterisation, mechanistic toxicity, exposure, and risk assessment data, enriched with atomistic, structural, molecular, and periodic table-based descriptors. Built on an adapted ChemBL schema, nanoPharos captures NMs' complexity from unit cell to macroscopic properties, linking rich bibliographic, provenance, and scientific metadata. Case studies demonstrate scalability for advanced materials, while integration with platforms like nanodash and Zenodo enhances FAIRness. Evaluation via Joint Research Centre maturity indicators shows strong compliance, with ongoing work towards full ontology integration and advanced API queries.

1. Introduction

Chemicals and advanced materials, including nanomaterials (NMs), are increasingly becoming an integral part of everyday life, being essential to health care, electronics, energy, transport, and housing materials, due to their novel and enhanced functional properties (*Why does the EU support research and innovation for chemicals and advanced materials?*, n.d.). Besides their multitude of benefits, advanced materials need to be safe, sustainable, and circular (*Why does the EU support research and innovation for chemicals and advanced materials?*, n.d.) and contribute towards the implementation of the European Green Deal (EU, n.d.), which binds the European Union (EU) to become climate neutral by 2050 through green growth and innovation (*Why does the EU support research and innovation for chemicals and advanced materials?*, n.d.). Furthermore, advanced materials offer the potential of substituting critical raw materials and contribute towards moving away from fossil fuels to other energy sources (EC, n.d.).

The development of novel and advanced materials, including NMs, is expected to bring positive socioeconomic impact from scientific research and its commercial applications. Nevertheless, the increased production and use of novel and complex materials leads to more complex production and handling processes, substantially increasing the lifecycle of a material. This can lead to increased risk of environmental releases and potential consequences such as undesirable environmental and biological effects (Lead et al., 2018). As a result, the risk and hazard assessment of novel materials requires a substantial effort, with current regulatory testing approaches relying heavily on animal testing, which raises ethical concerns, and incurs significant costs (Tsiros et al., 2022). For this reason, key industrial stakeholders published the Materials 2030 Manifesto, which called for the establishment of a strategic roadmap for the effective governance of advanced materials to ensure successful Research and Innovation (R&I) of applications of advanced materials (*Materials 2030 Manifesto: Systemic Approach of Advanced Materials for Prosperity – A 2030 Perspective*, 2022). Within the manifesto, the authors

* Corresponding authors: AGP and AA at NovaMechanics Ltd., Nicosia, Cyprus and IL at University of Birmingham, Birmingham, UK.
E-mail addresses: papadiamantis@novamechanics.com (A.G. Papadiamantis), i.lynch@bham.ac.uk (I. Lynch), afantitis@novamechanics.com (A. Afantitis).

https://doi.org/10.1016/j.impact.2025.100602
Received 28 August 2025; Received in revised form 26 November 2025; Accepted 27 November 2025
Available online 29 November 2025
2452-0748/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

called for uniting digital technologies, e.g., high performance computing, big data and Artificial Intelligence (AI), and material competences and capacities to "*revolutionise the digital modelling, simulation and screening of materials properties, materials development and production processes, unlocking the merge of computational and experimental material science*" (*Materials 2030 Manifesto: Systemic Approach of Advanced Materials for Prosperity – A 2030 Perspective*, 2022).

In their response, following extensive consultation with stakeholders, the EU acknowledged the importance for advancing the R&I and Research and Development (R&D) of advanced materials. The EU also emphasised the long timescale needed for the innovation process and noted the insufficient level of digitalisation currently as a barrier to progress. Using conventional methods, R&I and R&D processes, currently, require between 10 and 30 years to design and develop advanced materials and have them embedded into products on the market (EC, 2024). They also acknowledged that the development and exploitation of digital tools, like artificial intelligence (AI), can accelerate the discovery and production of innovative materials. The EU also concluded that the diverging approaches in data digitalisation, e.g., lack of harmonisation of data description (metadata) and data documentation formats and databases, hampered progress towards the necessary digital transition (EC, 2024).

Implementing the digitalisation of R&I in advanced materials requires a number of components and services, collectively referred to as the digital ecosystem. These include data infrastructures, digital modelling tools, common data analytics approaches utilising agreed vocabularies and ontologies, and accessible AI tools. These infrastructural components rely on large volumes of high-quality data, which will enable the "*analysis and interpretation of data from various characterisation techniques, improving modelling, and [..] suggesting composition or structure of new materials*" (EC, 2024).

This R&I approach requires data which have been produced in a transparent manner and are described sufficiently to allow assessment of their suitability for re-use for other purposes and under various contexts, even those for which they were not initially intended. Such data can feed Alternative Testing Strategies and Integrated Approaches to Testing and Assessment (IATA), which have been proposed to support and substitute in vivo experimental data in accordance with the 3Rs principles (replacement, reduction, and refinement of animal testing) (Tsiros et al., 2022). While IATAs include both in vitro and computational approaches, the digitisation of novel and advanced materials R&I and R&D processes (EC, 2024) requires high-quality structured data for the development of robust and validated computational approaches. In silico models can play a key role in predicting the properties, mechanism of action, and behaviour of materials in different environments and can be integrated into Safe and Sustainable by Design (SSbD) frameworks for the (re)design, improvement, and production of progressively "safer" materials.

To achieve this, curation of existing datasets from the currently fragmented, and in most cases inaccessible, nanosafety domain (and beyond) is required. These data need to be processed, cleaned, and FAIRified to maximise their value and allow combination with other similar datasets for more robust and validated analysis and model development. For this reason, research data management needs to follow the FAIR (Findable, Accessible, Interoperable, Reusable) data principles (Wilkinson et al., 2016), something which was also reported in the EC communication (EC, 2024). To maximise the data's added value, the same principles need to be applied to newly produced computational data. This is facilitated using modern data management tools for (meta)data capture, structuring, and harmonisation such as Electronic Laboratory Notebooks and protocol repositories, which move the FAIRification process to the earliest phases of the data lifecycle (Papadiamantis et al., 2020a; Martinez et al., 2020).

The FAIR Data Principles aim to guide data producers in making their data Findable, Accessible, Interoperable, and Reusable by other users and for diverse purposes (Wilkinson et al., 2016). It must be noted here that while the FAIR Principles were introduced to support both human and machine (remote computer access and data retrieval) actionability, in reality their implementation is mainly focussed on the latter as needed to facilitate the digital transition. Intentionally, the FAIR Data Principles do not explicitly provide guidance on technological implementations, or guidelines on the specific responsibilities and actions that the data producers need to follow to maximise the FAIR-ification level of their datasets. It is left to each respective "community" to decide on the means of implementation based on their data types, reuses, and existing norms and approaches, through a sociotechnical agreement. This has led to various attempts to interpret (Jacobsen et al., 2020; GoFAIR, n.d.) the FAIR data principles and provide a roadmap and examples for their implementation, and to "communities" developing practical guidelines for "their" data producers who often have limited FAIR-awareness and a non-technical (computer science) background (Papadiamantis et al., 2020a).

Considering the required technological implementations and domain-specific consensus (sociotechnical agreements) (Schultes, 2023), the GO FAIR Foundation interpretations (GoFAIR, n.d.) of the FAIR Data Principles are based on community consensus built via consultations with FAIR experts and domain specialists to "*ensure, as much as possible, interoperability, machine-actionability, global participation and convergence towards accessible, robust, widespread and consistent FAIR implementations*" (GoFAIR, n.d.). Nevertheless, the GO FAIR Foundation does not offer explicit technological solutions, but assists communities with re-scoping and reusing others approaches, or building 'their own domain-specific solutions.

To address this lack of publicly available high-quality structured and ready-for-modelling NMs and nanosafety datasets, we present here the current version of the nanoPharos database (*nanoPharos Database*, n.d.). nanoPharos contributes towards the digitalisation of R&I and advanced materials (including NMs) R&D, and offers users FAIR, Open, and high-quality ready-for-modelling (i.e., tabular, structured, harmonised) datasets that are directly importable into computational workflows. These can be experimentally or computationally produced and literature-curated nanosafety and nano-biological interactions data. The datasets contain physicochemical characterisation, exposure, interactions and transformations, hazard and risk assessment data, which can be further enriched, where possible, with a series of computationally or bibliographically derived structural, atomistic, periodic table-based, and molecular NMs descriptors (Papadiamantis et al., 2020b; Papadiamantis et al., 2021). The integration of both material and system descriptors can be used to study the mechanism of action of materials from the atomic level (of the e.g., NM) to biomolecule, organelle, cell, organism, population and ecosystem levels, and thus supports the development of Adverse Outcome Pathways (AOPs) (Ankley et al., 2009), as well as nanoinformatics, Artificial Intelligence (AI), and machine learning (ML) workflows. Besides presenting the current version of nanoPharos, we present its evolution over time, and demonstrate its utility as an increasingly FAIR-compliant data repository.

2. Background and related work

The current state-of-the-art regarding NMs-related databases and the respective implementation of the FAIR Data Principles (Wilkinson et al., 2016), demonstrates a clear path towards increasingly FAIR practices. Current databases implement persistent identifiers (PIDs), have improved their minimum (and rich) metadata requirements, and have begun facilitating cross-platform data exchange. However, current databases face challenges in terms of expanding or adapting to the requirements and interpretations of the FAIR Data Principles (Jacobsen et al., 2020; GoFAIR, n.d.). These can be technical, organisational, or financial due to the fact that many of these databases were developed using public funding that may or may not be continuously available. Furthermore, most databases have focussed on indexing datasets without the specialised focus of harmonising the data output format to

ensure that it is machine-actionable and ready for direct import into computational workflows.

2.1. FAIR data principles and the GO FAIR foundation interpretations

The FAIR Data Principles (Wilkinson et al., 2016) provide guidance on how to make data and especially metadata, i.e., data about a dataset, findable, accessible, interoperable, and reusable. While the FAIR data principles are, in theory, aimed towards human and machine actionability, in reality a deeper study reveals their more technical nature and the fact that they lean heavily towards computer-computer actionability regarding digital resources. Despite their technical nature, and as described by Jacobsen et al. (2020), the FAIR principles do not offer specific technological solutions, rather, implementation is left to the individual "communities" (Jacobsen et al., 2020).

At the same time, a clear definition of a FAIR community does not exist and in practice can be anything from a single researcher, to laboratory group, a national or international project, or a wider scientific community, e.g., the Elixir community of life sciences researchers. As a result, the FAIR principles can be interpreted and implemented differently and inconsistently, even in cases of very similar, in terms of scientific content, research communities. This can lead to incompatible solutions, which hinders true data interoperability. To overcome this, attempts have been made to articulate high-level interpretations and implementation considerations by FAIR experts, e.g., the GO FAIR Foundation whose interpretations (GoFAIR, n.d.) are based on the work of Jacobsen et al. (2020), to assist with wider participation and facilitate convergence towards adaptable technological solutions (Jacobsen et al., 2020; GoFAIR, n.d.).

Based on the GO FAIR Foundation interpretations, data and respective metadata are stored as separate digital records. The metadata record is based on a clearly defined schema, i.e., a specification that defines the required metadata fields and how they describe the data they refer to (FAIR Data Principle F2). For data and metadata (henceforth (meta) data) findability also relies on the use of distinct Globally Unique Persistent and Resolvable Identifiers (GUPRIs) for both (meta)data (FAIR Data Principle F1). These GUPRIs are then mutually referenced, i. e., the dataset references the metadata record GUPRI and vice versa (FAIR Data Principle F3) and each record is indexed in an appropriate registry, which is searchable from both humans and machines (FAIR Data Principle F4).

Accessibility, as per the GO FAIR Foundation interpretations (GoFAIR, n.d.), is based on the use of clearly defined communication protocols (FAIR Data Principle A1), e.g., Application Programming Interfaces (APIs), which is freely and openly accessible and universally implementable (FAIR Data Principle A1.1). Furthermore, the protocol used needs to implement a clear authentication and authorisation process where needed (FAIR Data Principle A1.2), e.g., using Single Sign-On (SSO) services with individual credentials and respective permissions. One key aspect of metadata accessibility is a clearly defined longevity plan (FAIR Data Principle A2). This interpretation is based on the realisation that data can become inaccessible over time, due to policy and regulation issues, e.g., human sensitive data, or by accident, e.g., corrupted files. If these data have been used under certain contexts, e.g., in publications, for model development, and have been respectively referenced, this lack of accessibility can impact the transparency and validity of the work presented. For this reason, a separate metadata record describing these data is envisaged, which is stored separately and is available through a registry that will remain functional over the long term to describe the data's nature and provenance (GoFAIR, n.d.) even when the data itself are no longer accessible.

To promote interoperability, the GO FAIR Foundation interpretation (GoFAIR, n.d.) requires that (meta)data records are linked to clearly defined structured vocabularies or ontologies (FAIR Data Principle I1), which themselves follow the FAIR Data Principles (FAIR Data Principle I2). This means, that key terms and concepts in the (meta)data records have been annotated using clearly defined vocabulary or ontology terms, which are described via GUPRIs and are machine-actionable. Each term in the dataset is thus clearly defined, including each relevant relationship with other existing entities (GoFAIR, n.d.). In this way, when queried, especially through machine actionability, a common unambiguous understanding of the digital resource is achieved. FAIR Data Principle I3 foresees that the metadata records include qualified references to other metadata. This, for example, can include versioning of a dataset or of the metadata record, in case changes are made to the original files, which leads to the publication of a new file with a new GUPRI. In this case, any previous versions should be referenced along with a description of the respective changes. Another example includes referencing the original resources, through e.g., Digital Object Identifiers (DOIs), that were used to develop a literature curated dataset for meta-analysis purposes.

Finally, the GO FAIR Foundation interpretations for reusability (GoFAIR, n.d.) include the provision of the (meta)data being described with rich metadata (FAIR Data Principle R1) that allow a new potential user to decide whether they can reuse the retrieved data for their intended use. In practice, this means providing enough scientific metadata to fully explain the data, how they were produced, for what purpose, and whether it is possible to combine these with other data to create a larger dataset. FAIR Data Principle R1 differentiates from F2 in the sense that the latter is used for findability purposes, i.e., attribute-based search and query (GoFAIR, n.d.), while R1 refers to the nature (purpose) of the data in question. Under that guise, FAIR Data Principle R1.1 requires a clear (re)usage license for the (meta)data records, e.g., CC-BY-4.0, so that the user knows the conditions of access and reuse of the data. FAIR Data Principle R1.2 requires documentation of the provenance information related to the dataset, i.e., who produced and owns the data, the methods used to produce the data, and the funding under which the data were produced (GoFAIR, n.d.). Finally FAIR Data Principle R1.3 aims to ensure that the published (meta)data records follow established and agreed upon community standards that are expressed through the FAIR data principles and a set of minimum scientific information that allows for the (meta)data to become fully understandable, e.g., as per Chetwynd et al. (2019) regarding the minimum information about Nanomaterial Biocorona Experiments (MINBE) (Chetwynd et al., 2019).

We note here that one key obstacle to the full implementation of the FAIR Data Principles, is the lack of globally acceptable metadata schemas for nanosafety and nanotechnology research. Several attempts have been made to define such standards to promote (meta)data convergence and interoperability within the scientific community. Papadiamantis et al. (2020) promoted a set of Scientific FAIR Data Principles to assist data producers on the actions required on their part to maximise, as much as possible, the FAIRness of their nanomaterials (NMs) safety data (Papadiamantis et al., 2020a). These were accompanied with case studies on the essential information reporting on NMs agglomeration as a source of in vitro delivered dose variations critical to human hazard assessment and achieving consensus on terminology and metadata usage regarding NMs dissolution (Papadiamantis et al., 2020a). Similarly, Erkimbaev et al. (Erkimbaev et al., 2015) and Elberskirch et al. (Elberskirch et al., 2022) have proposed metadata schemas for reporting the characterisation of NMs and for NMs exposure experiments and analysis, respectively. Nevertheless, none of those have been widely adopted as yet, in part because they have not yet been made machine-actionable, given user-friendly web interfaces to facilitate annotation of datasets or embedded into key data repositories to enable streamlined data FAIRification workflows.

Recently, Exner et al. (Exner et al., 2023) highlighted the need for the development of metadata standards that are widely accepted and act as the basis for (meta)data machine actionability and promote FAIRification. In parallel, Punz et al. (Punz et al., 2025) built on the concept of the instance maps, originally developed via the NMs informatics knowledge commons (NIKC) platform (Amos et al., 2024), to promote a

tool that acts as a form of schema for the visualisation of experimental practices and materials and data provenance to bridge the gap between the theoretical nature of the FAIR Data Principles and the everyday experimental practice. Finally, for assessment of the FAIRification level of datasets, and respective metadata, Ammar et al. developed the NanoSafety Data Reusability Assessment (NSDRA) framework, which is a machine-actionable representation of 12 minimum reporting standards using FAIR maturity indicators for the annotation of datasets as a means to assess their FAIRness-level (Ammar et al., 2024). Finally, the EU's Joint Research Centre (JRC) published the JRC FAIR Data Guidelines, as part of the JRC Data Strategy, to support and enhance the FAIRness of JRC data (Lowenthal et al., 2025). The guidelines propose 5 progressive levels of FAIR Maturity, namely FAIR start, FAIR play, FAIR go, FAIR share, and FAIRest of them all, based on specific FAIR maturity indicators. It is accompanied by a FAIR maturity assessment grid that data re-users can use to evaluate the FAIR maturity of their dataset or offered resource, e.g., registry services.

Here, we present the current version of the nanoPharos database, which is the most updated version containing the accommodations that were developed within the three collaborative cases studies (with the DIAGONAL, CompSafeNano, and INSIGHT projects) described in this work. We also describe the philosophy behind the design and implementation of nanoPharos, which has evolved to accommodate metadata describing increasingly complex NMs and advanced materials datasets. Using the 3 case studies, the increasing FAIRness and evolution of nanoPharos is also demonstrated.

3. Methods

3.1. Underlying model and database structure

3.1.1. Conceptual data model

To maximise the information, analytical potential, and added value of the nanoPharos datasets, a detailed data model has been developed

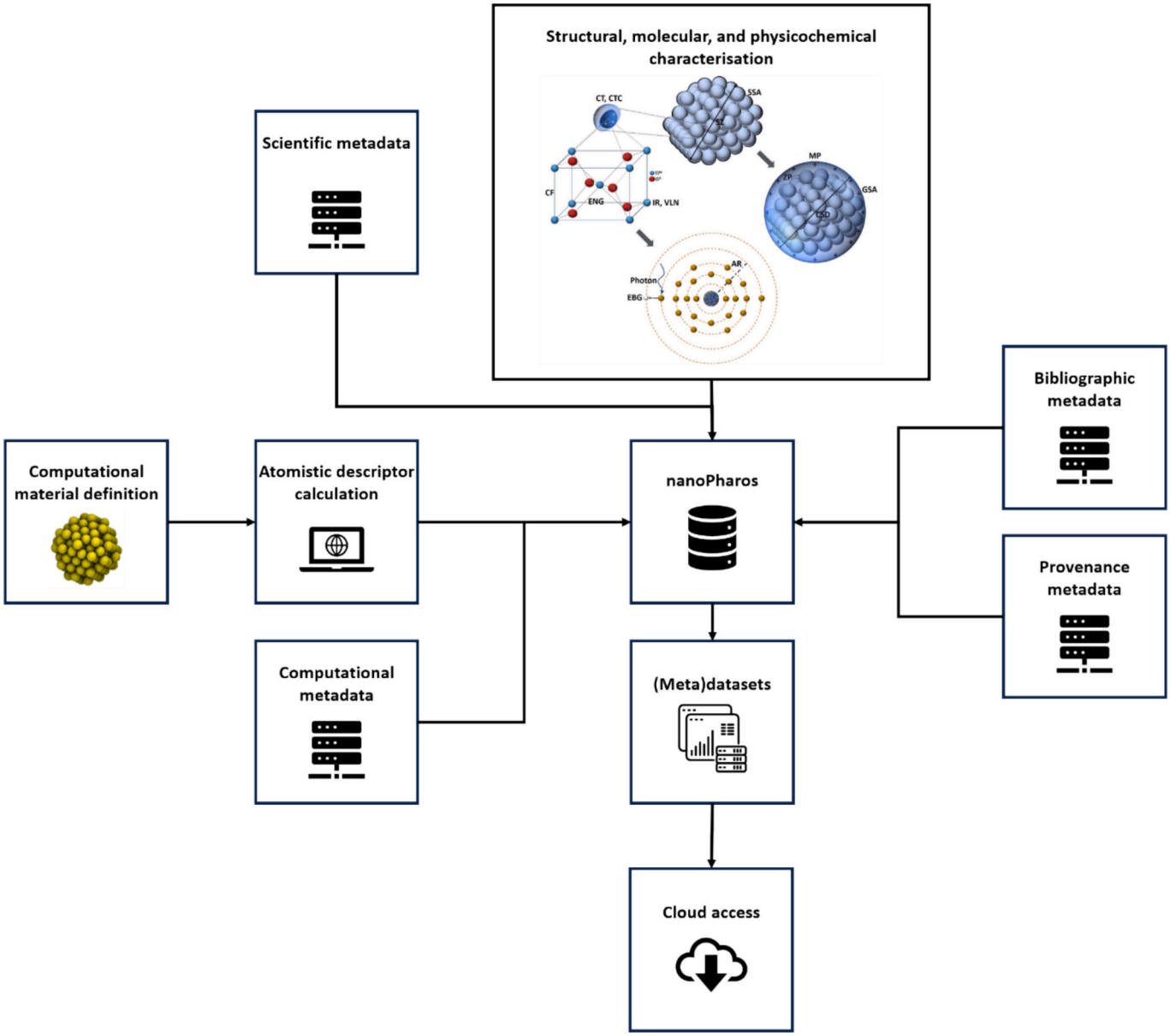

Fig. 1. Schematic representation of the nanoPharos underlying model and how the data are connected to the respective scientific, bibliographic, and provenance metadata illustrated using a TiO_2 NM. The abbreviations presented in the 'Structural, molecular, and physicochemical characterisation' box read: AR: Atomic Radius, CF: Chemical Formula, CSD: Corresponding Sphere Diameter, CT: Coating, CTC: Coating Charge, EBG: Energy Band Gap, ENG: Electronegativity, GSA: Geometric Surface Area, IR: Ionic Radius, MP: Morphology, SSA: Specific Surface Area, SZ: Size, VLN: Valency, ZP: Zeta Potential.

(see schematic representation in Fig. 1). The original data model was built based on the ChemBL schema (*ChemBL Database Schema*, n.d.), which has now been adapted and expanded to accommodate the specialised and evolving characteristics of NMs and nanosafety data (for a more comprehensive analysis of the adaptations, please see Fig. S1 and respective text in the "ChemBL to nanoPharos adaptation and expansion" Section of the Supplementary Information). The data model describes the links and dependencies between the different types of data included in nanoPharos along with their respective metadata. The nanoPharos data model is based on a detailed description of a NM, starting from the atomic level, i.e., unit cell and crystallography information, and builds up to macroscopic information, i.e., physicochemical and surface chemistry characterisation.

The effects of NMs are also described using different data types, e.g., in vitro toxicity targeting a range of endpoints (cytotoxicity, genotoxicity, immunotoxicity, developmental toxicity, etc.), AOPs, and ecotoxicity (e.g., NMs effects on daphnids, earthworms, fish etc.). The exposure and release routes, e.g., release to environmental compartments, binding to soils, road dust etc., are catalogued as part of the metadata to facilitate data provenance and context, FAIRification, and the development of meaningful metadata records. At the same time, these metadata are also available as covariates for meta-analysis and data-driven modelling when combining data from heterogeneous studies.

This approach reflects the reality that the border between data and metadata is fluid, and that the metadata of one study can become the data of another, and thus allows the translation of the captured (meta) data into Scientific Knowledge Graphs, along with the data it describes, which allows for the visualisation of data flows and interconnections, defines the required metadata, and enables exploration and understanding of the connections between data originating from diverse sources (Hogan et al., 2021). Users are then able to better understand the nanoPharos data model and how a specific NM, organism, endpoint, or other factor has been used in different studies and its impact on the respective outcomes, without having to rely on complex schema documentation.

One of the key implementations is the need to address the complexity of NMs and especially the cross-batch variation, due to the stochastic nature of synthesis, that can be observed in commercially available materials. This variation can be compounded by differences in storage and exposure to different environments leading to different transformations (Svendsen et al., 2020; Johnston et al., 2020). nanoPharos manages this by defining different batches or altered forms of a "parent" NM, i.e., the original as-produced NM before any characterisation or assays have been performed using it. The lifecycle of these batches or transformed materials are monitored using the European Materials Registry (ERM) identifier, which is one of the globally unique identifiers used to define a material (van Rijn et al., 2022). Other identifiers used include the Chemical Abstracts Service (CAS) number, EC number, International Uniform Chemical Information Database (IUCLID), Simplified Molecular Input Line Entry System (SMILES) representation, and the extension of the InChI chemical identifier to NMs, via the NInChI which describes the "parent" or ideal composition of a material (Lynch et al., 2020).

The structural information of a material can be based either on bibliographical data, e.g., from the Materials Project website containing information on 178,627 materials (*The Materials Project*, n.d.), or on experimental data if researchers use high-quality XRD data and Rietveld refinement to recreate their NM's unit cell and calculate the respective information (Ellis et al., 2018). At the same time, computational representations of specific materials have been developed, which can be used for the calculation of atomistic descriptors using simulations and physics-based modelling to enrich experimental datasets. Thus, molecular and periodic table-based descriptors have been included in nanoPharos, including metal electronegativity, atomic and ionic radius etc. This has led to a total of 119 materials-related descriptors that can be used to describe a material. These comprise of 36 physicochemical descriptors (including 18 that can be extracted from TEM images using the NanoXtract tool (Varsou et al., 2020; NovaMechanics, 2020)), 3 structural descriptors, 18 molecular and periodic table-based descriptors and 62 atomistic descriptors (see Table S1 in the Supplementary information for a full list).

The exposure, interactions, hazard and risk assessment part of nanoPharos allows users to upload and import relevant data along with methodological assay-related information as part of the metadata. In this case, metadata are also handled as data points since methodological assay information have been found to play a statistically significant role during meta-analyses (Labouta et al., 2019) and ML model development (Papadiamantis et al., 2020b), especially in cases where data from different sources and different methodological approaches have been combined. For this reason, nanoPharos has been designed to enable methodological fields to be used as variables or covariates during meta-analyses and ML development across heterogeneous studies, e.g., the cytotoxicity assay type as different cellular toxicity assays measure different aspects of cell death and thus assay differences may explain more of the variability in a dataset than the fact that they utilised different sized NMs, for example. Thus, nanoPharos offers users the opportunity to import data on the biological entities and/or biological entities categories tested and the assay type, along with respective information on the experimental setup and end-points used.

As seen in Fig. 1, the data imported into nanoPharos are directly linked to their respective metadata. In this way, each data point is linked to information describing how it was produced, who produced and curated that data, who owns the data, the license under which a dataset is available, any relevant publications, and more. nanoPharos accommodates 3 types of metadata to satisfy the FAIR Data Principles requirements for rich metadata for findability (FAIR Data Principle F2) and reusability (FAIR Data Principle R1):

- Bibliographic metadata: dataset title, description, owner(s), data producer(s), curator(s), contact detail(s), relevant publications, unique IDs and descriptors, e.g., DOI, ORCID, file type and size.
- Provenance metadata: methods used to produce the data, date of data production, date of modification (where applicable), versioning.
- Scientific metadata: protocols, methods, instruments used, analytical and computational algorithms, software used and versions.

In this way, a relational database schema was developed, which is compliant with the FAIR Data principles, and that maps the complex relationships and dependencies in nanosafety, and materials safety, data. Key elements of the schema include entity-relationship diagrams, precise data types, constraints to ensure data integrity, and indexes for optimised query performance. This approach allows for the development of a scalable database, the data model of which can be extended based on specific requirements leading to the integration of different data types and respective metadata management.

3.1.2. Logical/physical design

The current management system of nanoPharos is based on MySQL. The benefit of using MySQL over other solutions, e.g., NoSQL, is that it can better handle predefined schemas, like that of nanoPharos, for the management of structured data used for the development of data-driven workflows. NoSQL, on the other hand, is better at handling unstructured or dynamic data. Another advantage of MySQL is that it is table-based, i.e., can better accommodate scientific datasets, while NoSQL is better for documents and graph databases.

MySQL provides databases with increased performance and scalability, which leads to better web-based performance regarding data retrieval and transaction rate speeds thus streamlining the management

of multiple simultaneous queries. As an open-source system, MySQL is a cost-efficient solution allowing for better budget allocation towards more complex feature and tool development. MySQL is easy to setup, maintain and customise, and it is compatible with multiple operating systems and platforms, and has extensive community support making it easy to debug and ensuring a high stability and security level can be maintained, which is particularly important when handling sensitive data. MySQL provides readily available features like encrypted connections and data-at-rest encryption options.

Database development employed a set of tools and technologies that ensure functionality robustness, scalability, and back and forward compatibility, as well as industry-standard database performance. Database development has been based on Java, JavaScript, and HTML5. Java has been used for backend development, as it offers robust and platform-independent features, while its API allows the streamlined handling of complex tasks and operations. (Baddam et al., 2018) The nanoPharos backend system management is supported by Wildfly, which is a scalable and modular open-source application server that offers quick initialisation and efficient resource management for high performance and responsiveness.

On the frontend, JavaScript and HTML5 offer a flexible and friendly user-experience through real-time user interface (UI) interactions. These are enhanced by CSS3, which improves the user experience using advanced styling, aesthetic, and user-friendly solutions in the graphical user interface (GUI), without loss in functionality. Synchronization of the backend with the frontend, for a seamless user experience, is performed using the ZK Framework. The ZK Framework is used to manage AJAX (Asynchronous JavaScript and XML) and JavaScript operations during UI development and is used during the development of Rich Internet Applications (Ochoa-Zezzatti et al., 2009). The use of the ZK Framework reduces complexity and ensures a smooth and reliable user experience and an advanced and fully responsive GUI design.

The ZK Framework enables Rich Internet Applications using a server-centric model that offloads complex data processing to the server side, reducing client-side scripting. Its extensive suite of AJAX components automatically handles UI updates for dynamic, real-time interfaces, while the Model-View-ViewModel (MVVM) design pattern separates data logic from the UI to enhance maintainability and scalability. Additionally, ZK's event-driven programming paradigm further streamlines application responsiveness by simplifying the synchronization of UI interactions. Integrating CSS with the ZK Framework, nanoPharos offers a user-friendly interface that balances functionality and aesthetics.

The data security and access control of nanoPharos is managed using Keycloakto set up authentication and authorisation protocols. The identity and access management processes support role and privileges assignment, so that different users, e.g., database managers, curators, data re-users, are offered different functionalities and views. Keycloak provides increased security through user authentication and supports SSO and user federation functions. Keycloak ensures sensitive data protection, in accordance with the General Data Protection Regulation (GDPR), restricting access to users based on their assigned roles and rights.

3.1.3. User interface

Based on the design and development described in Sections 3.1.1 and 3.1.2 and the nanoPharos data model, the nanoPharos UI was designed to be task-oriented, role-specific, and aligned to the underlying database schema and its focus on delivering ready-for-modelling datasets. For this reason, the design principles of the UI are focussed on the development of customised, case-specific datasets by choosing the materials and/or assay descriptors, i.e., data and metadata, of interest for specific materials from those included and available in the database.

The landing page of nanoPharos (Fig. 2) guides users to the advanced configuration section, and not the general keyword search, where they can choose from the available (meta)data tables, attributes, and values. A side panel displays clickable boxes for selecting tables, while dropdowns, radio buttons, and an output configuration option allow users to define and refine their searches according to logical operators, equality, or inequality. In this way, users can include or exclude attributes, move them between lists, and apply additional filters. Furthermore, dedicated search boxes enhance navigation, and the ability to click on radio

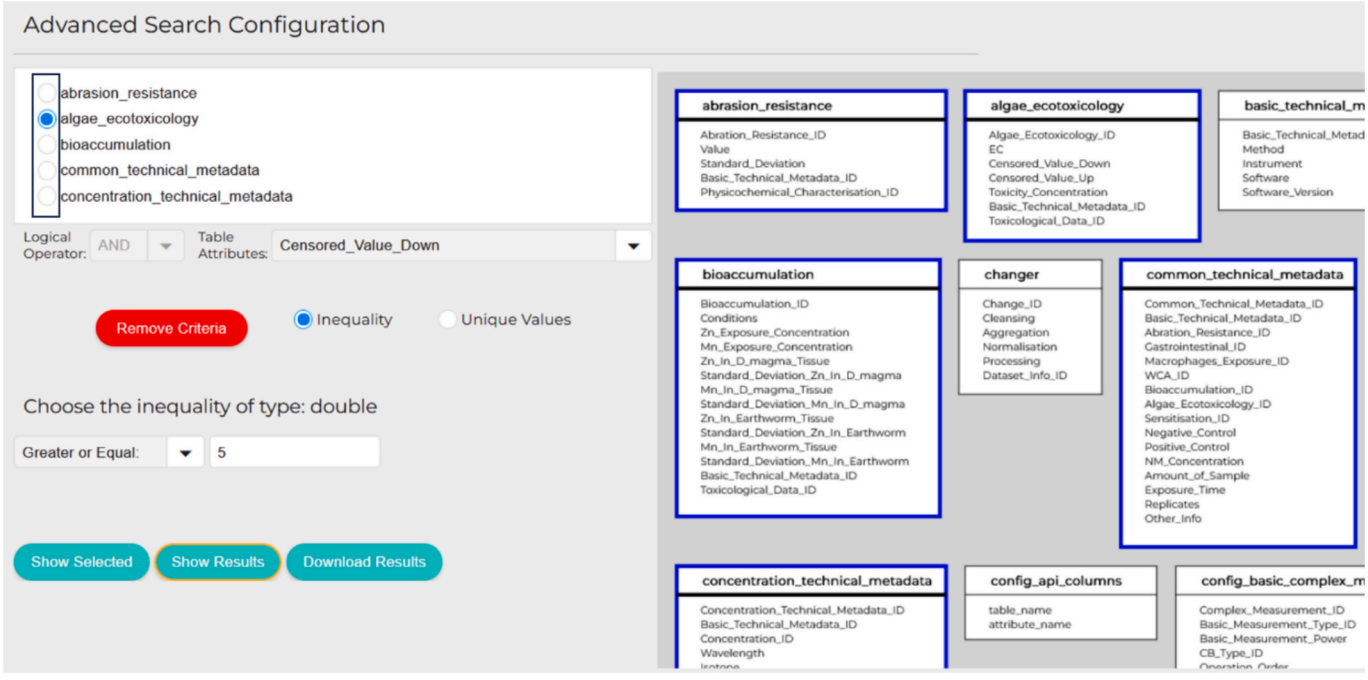

Fig. 2. In the Advanced Search Configuration Page users can pick the data and respective tables of interest from the complete set in the database. Users can choose multiple tables and switch between them using radio buttons (see top left black rectangle). Users can also define and refine their searches according to logical operators, equality, or inequality (bottom left). In the example shown here, the user has selected to include data on algae ecotoxicity, bioaccumulation, and abrasion resistance, including respective technical and NM-related concentration metadata.

buttons to choose inequality options, e.g., “Greater” or “Lesser” or “Equal”, enables data retrieval based on specific user needs. The system then exports the refined results in a clearly presented tabular format, supporting efficient and targeted data exploration and direct import into computational workflows.

As stated earlier, nanoPharos has provisions for embargoed and sensitive data, which may be locked and is thus inaccessible to general users. For this, Keycloak is being employed and specific permissions and privileges are applied on a case-by-case basis. When logging in to access sensitive data, users are redirected to the nanoPharos database management system, where 4 options are available (where applicable based on specific privileges):

- Embargoed/Locked or Project-specific Data
- Bulk Insert
- Single Insert/Modify
- Table Configuration

Users that belong to a specific project are able to see data that may be under embargo, i.e., to be released based on data owner exploitation or following expiration of the embargo period, or data that are locked due to sensitivity, e.g., personal data, commercially sensitive data. The data that are visible to specific users is defined based on rules defined by each individual project or contributor and are addressed individually.

For data curation, the Bulk Insert option (Fig. 3) allows data providers / database curators to upload data using templates developed for each parameter. The users use a dropdown menu to pick the metric they want to upload (Fig. 3 top) and the system provides them with the option to download the preformatted template in CSV (Comma Separated Values) format. The users fill in the template, upload it, and the system automatically links the template columns with the respective database ones (Fig. 3 bottom). To assist users with the curation process, a guide has been developed (NovaMechanics Ltd, 2025) and included in the Bulk Operations Menu panel.

Curators can use their own templates also, but they will then have to manually map the CSV file they upload to the required database attributes (Fig. 4). In this case, the curators would see the image presented in Fig. 4 (top) and would need to click on the Select button (see black rectangle under Edit) for the respective row and chose the matching column from their uploaded template (Fig. 4 middle), e.g., CAS Number in this example is mapped to CAS, which is required in nanoPharos, and press the Add and then confirm buttons (see black rectangle in Fig. 4 middle). Repeating the procedure for all relevant columns, curators can quickly and easily map an entire template to nanoPharos (Fig. 4 bottom) and upload it.

The users will also need to define, where applicable, the unit(s) of their data. These will automatically be converted to the default units used in the database for the specific descriptor, if needed. When clicking upload, the system automatically checks the dataset for validity and confirms that it meets the required (meta)data types, dependencies, and requirements. If the test is passed, the dataset is uploaded and automatically published in nanoPharos based on the choices made by the data owners and implemented by the curators (these roles may be undertaken by the same or different person(s)).

To add or modify single entries (Fig. 5), the Single Insert/Modify menu offers curators the ability to choose a specific table and perform the desired action. They can manually add attributes, while the system provides information on the data type required for each field. Built-in quality control scripts verify correct formatting and check for missing mandatory fields. To modify data, curators pick a table, select attributes to update, and retrieve existing values for specific entry IDs, which can then be edited or removed. Each successful action triggers a confirmation message, helping maintain transparent and reliable record-keeping throughout the curation process. Similar to the Bulk insert option, the nanoPharos Curation Guide (NovaMechanics Ltd, 2025) offers users with guidance on the single entry addition or modification and the link has been included in the respective Single Insert/Modify panel.

Database managers and administrators can access the Table Configuration and Table Attribute Configuration pages (Fig. 6). The Table Configuration feature (Fig. 5, top) lets database managers or administrators control how each table appears and the required information. Through this functionality, the tables can be renamed, its

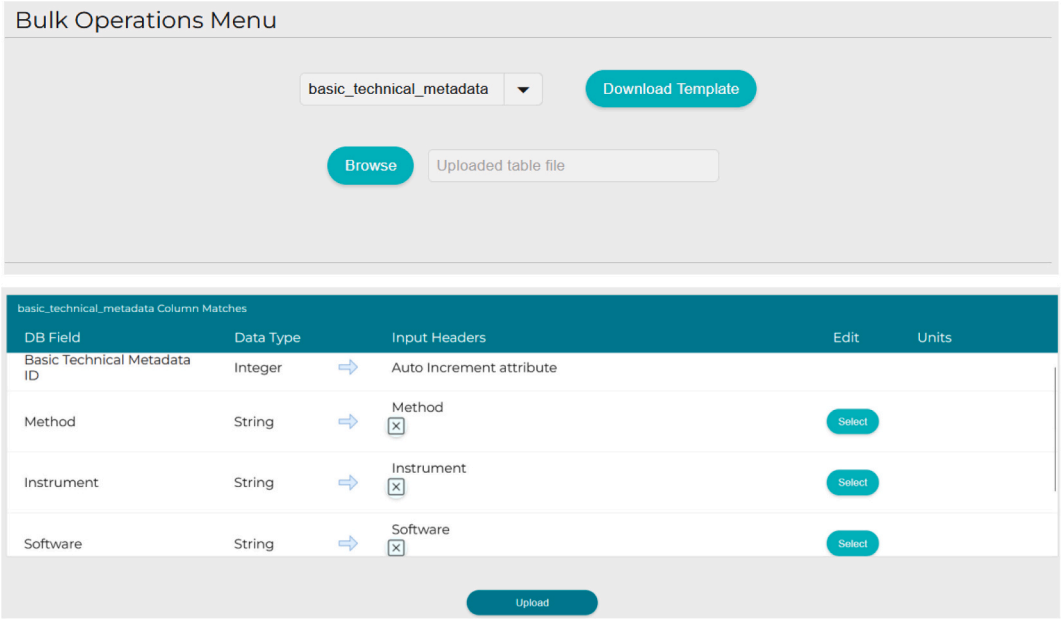

Fig. 3. Bulk Insert Data menu page (top) and Data Upload and validation (bottom). Bulk data insert allows users to import full datasets automatically based on predefined templates provided by the database. Following upload, the system automatically maps the template to the database attributes. If using 3rd party templates, users will need to manually map their template to the nanoPharos attributes. Users can also define, where applicable, the unit(s) of their data, which are then automatically converted, if needed, to the default used by the database. The system validates the dataset and if all checks are passed it is uploaded to the database. A red star denotes mandatory or automatically filled fields. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

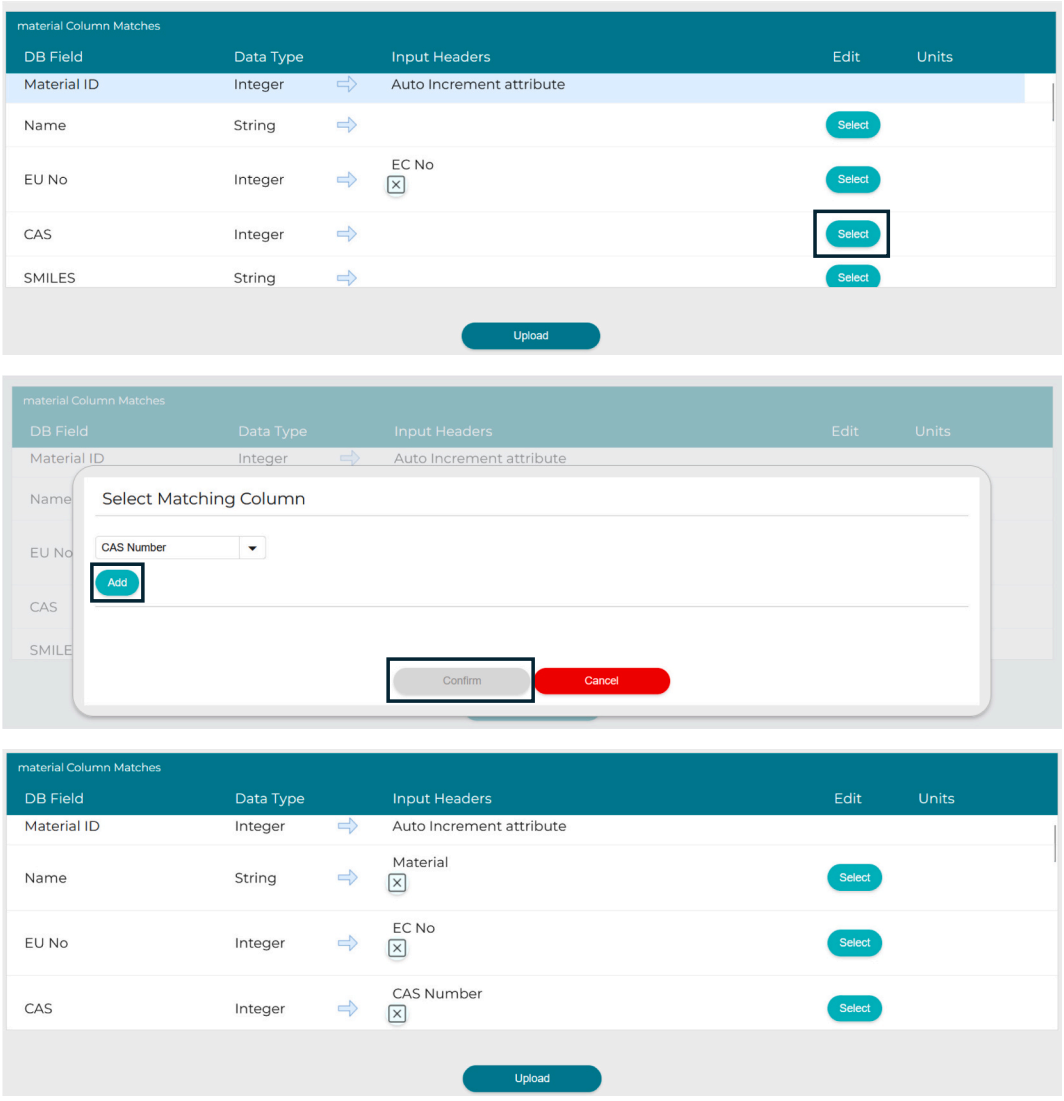

Fig. 4. Users using their own templates (instead of those provided by nanoPharos) will need to manually map the columns in their templates to the database-specific attributes before uploading their dataset to the database.

visibility changed (e.g., to decrease findability or remove if it is not required in a specific database instance), and descriptions added to aid Findability. Security features prevent duplicate names to avoid confusion. Through the Change Table Attributes option (Fig. 6, bottom), table attributes can be renamed or changed (e.g., for technical or developing purposes), their visibility defined, and descriptive details or indicative values can be provided. When any changes made through these pages are implemented, any other components in the specific dataset / database instance containing these tables, e.g., Advanced Search and Output Configuration, are automatically updated to maintain a consistent, bug-free, and user-friendly environment.

3.2. Methodology for FAIRification

nanoPharos is a data registry, which is increasingly compliant with the FAIR Data Principles (Wilkinson et al., 2016) and applies the Scientific FAIR Data Principles (Papadiamantis et al., 2020a) to facilitate use by non-technical data producers. The FAIRness implementation underpinning nanoPharos is based on the GO FAIR Foundation interpretation of the FAIR Data Principles (GoFAIR, n.d.). To monitor the nanoPharos FAIRness degree, we have developed a FAIR Implementation Profile (FIP) (Lynch et al., 2025), using the FIP Wizard (*FIP Wizard*, n.d.), where the respective FAIR implementation choices are catalogued. A FIP presents the choices made by a specific FAIR Implementation Community (FIC) for maximising the FAIRness level of their (meta)data (Schultes et al., 2020). Comparison of the nanoPharos FIP with those of other FAIR Enabling Resources (FERs) or FICs can be achieved via a FAIR Convergence Matrix (Sustkova et al., 2020). This matrix can be used to track the FAIR landscape between similar FICs, identify opportunities for optimisation around reuse and interoperation, and promote harmonisation between the FICs (Schultes et al., 2020).

As nanoPharos is designed to deliver high-quality datasets to promote the digitisation of materials R&I and R&D processes, it benchmarks itself based on the EU requirements for (meta)data FAIRification. For this reason, the FAIRness maturity level of nanoPharos is tracked using the JRC FAIR maturity questionnaire (the latest version of the nanoPharos FAIR maturity questionnaire is provided in the Supplementary information) (Lowenthal et al., 2025). Based on this, nanoPharos can be considered as having a high degree of FAIRness, although there are still gaps to overcome. The major current gap is the lack of ontologies or structured vocabularies implementation to support seamless machine actionability (FAIR Data Principle I1).

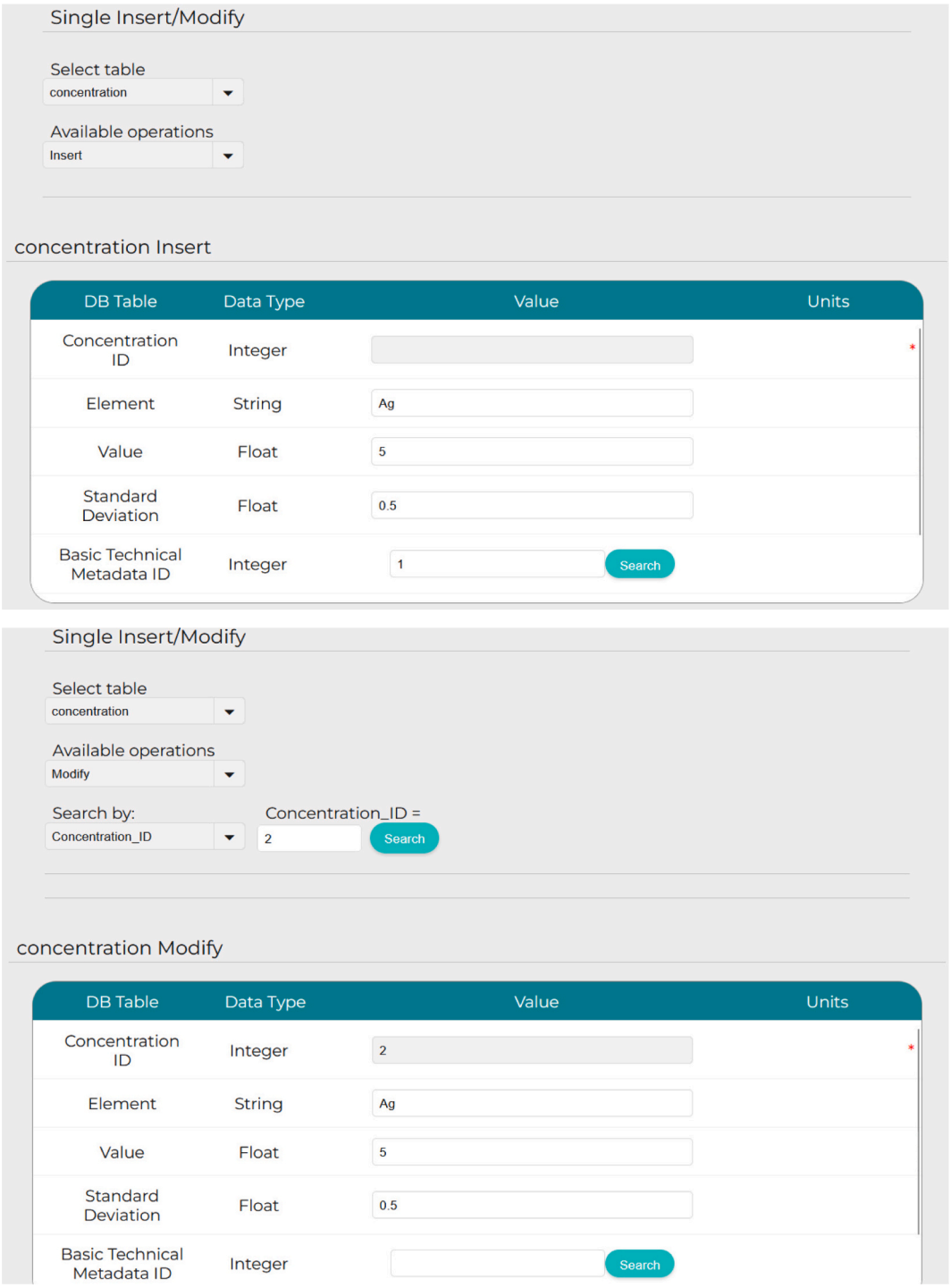

Fig. 5. The single insert (top) and modify (bottom) pages allow curators to manually input or modify existing (meta)data. The system provides information on the allowed data type for each field. Required fields are marked with a red star. Full documentation of changes is recorded for versioning and provenance tracking. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.1. Findability

Findability of the nanoPharos data is based on the assignment of unique identifiers (FAIR Data Principle F1) based on the RFC 3986 IETF standard for Uniform Resource Identifiers (URIs) (*RFC 3986: Uniform Resource Identifier (URI)*, n.d.). Similarly, metadata are published as a nanopublication using nanodash (FAIR Data Principle F4) (*nanodash*, n.d.). In this way, the metadata record becomes a digital resource, receives a GUPRI (FAIR Data Principle F1), and becomes human and machine findable via searchable indexes like nanodash and Zenodo (FAIR Data Principle F4). (Meta)data publications contain the others GUPRI referencing each other (FAIR Data Principle F3). For each dataset published in nanoPharos, a respective publication is created in Zenodo (*Zenodo*, n.d.) containing the dataset's title, summary, GUPRIs of (meta)data, license, and contributors. Through Zenodo, the datasets are also indexed in OpenAIRE (*OpenAIRE*, n.d.). In this way, the (meta)data are indexed in various searchable resources, i.e., nanoPharos, nanodash, Zenodo, and OpenAIRE (FAIR Data Principle F4).

Data findability is supported using rich metadata (FAIR Data

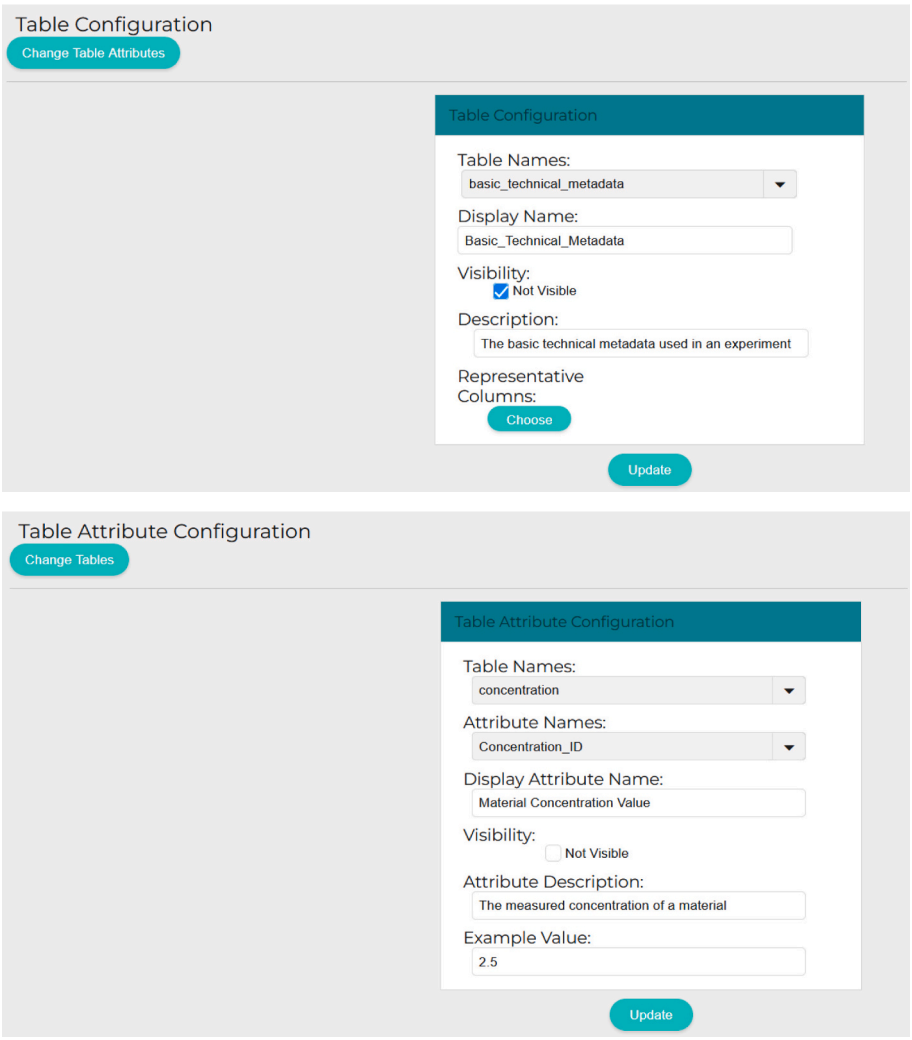

Fig. 6. The Table Configuration (top) and the Table Attribute Configuration (bottom) pages. Database managers and administrators can use these pages to update, rename, and add descriptions to specific tables or to attributes within tables. Note that such changes are not anticipated to happen frequently.

Principle F1) from the three categories described in the Conceptual Data Model section, i.e., bibliographic, provenance, and scientific metadata. For Findability the definition of metadata corresponds to descriptive metadata that allows humans and machines to identify a resource of interest through querying, searching, or filtering (GoFAIR, n.d.). This means that datasets are accompanied by information like title, summary or abstract, creator, owner, license, unique identifiers, publication data, publisher, keywords, etc. Defining "rich" for metadata is not a straightforward process and is usually left to the respective scientific communities to define it. For nanoPharos, "rich", for the purposes of FAIR Data Principle F2, corresponds to the datasets being described with the following bibliographic and provenance metadata:

- Unique identifier of the dataset in the form of a URI or DOI
- Title
- Summary/Abstract
- Data creators/producers, curators, owners (identified via ORCID IDs or another personal unique identifier)
- Access license
- Date created and published
- Dataset version
- File type and size
- Relevant publications describing or utilising the dataset (identified using DOIs or other unique ID)

The (meta)data, including the scientific metadata that fall mainly under FAIR Data Principle R1, are linked based on the nanoPharos underlying schema, which is based on that of the ChemBL database (*ChemBL Database Schema*, n.d.). This schema is expanded and customised to accommodate the specificities of nanomaterials and advanced and novel materials data (FAIR Data Principle F2).

3.2.2. *Accessibility*

Accessibility in nanoPharos is offered either through the web-human interface (*nanoPharos Database*, n.d.), or remotely using a Representational State Transfer (REST) API (*nanoPharos API*, n.d.). The REST API offers users the ability to retrieve datasets using their database identifier, as well as enabling programmatic interaction with other databases or modelling tools (FAIR Data Principle A1). As per the FAIR Data Principle A1.1 the API is open, freely accessible, and universally implementable (*nanoPharos API*, n.d.), although it does not currently support authentication and authorisation procedures (FAIR Data Principle A1.2) to access, for example, sensitive or embargoed data. The API also does not, currently, support retrieval based on keywords and there is no integration of search based on ontological terms.

For metadata longevity (FAIR Data Principle A2), we have opted to publish machine-actionable metadata records in nanodash (*nanodash*, n.d.) and Zenodo (*Zenodo*, n.d.), as described in the Findability section. Publication in these platforms ensures that even if the data are no longer available through nanoPharos, the metadata records will remain for the

long-term. Thus, if a nanoPharos dataset has been used and referenced in a published resource, the metadata record will persist and provide a description to sufficiently understand the data's nature, content, and provenance (GoFAIR, n.d.).

3.2.3. Interoperability

FAIR Data Principle I1 requires that the (meta)data are annotated using established structured ontologies or vocabularies. Currently, the datasets in nanoPharos do not support ontological annotation. Nevertheless, the nanoPharos datasets are available in machine-actionable forms like JSON and XML.

On the other hand, nanodash requires the use of annotated terms. As a result, the nanodash metadata publications are compliant with the FAIR Data Principle I1, as all terms are annotated using established ontologies like the eNanoMapper Ontology (eNMO) (Hastings et al., 2015), the Nanoparticle Ontology (NPO) (Thomas et al., 2011), Chemical Entities of Biological Interest (ChEBI) (de Matos et al., 2010), and more, which follow the FAIR Data Principles (FAIR Data Principle I2). Furthermore, the metadata can be also accessed in other machine-actionable forms like TriG, JSON-LD, N-Quads, and XML (FAIR data principle I1).

FAIR Data Principle I3 requires that (meta)data include qualified references to other (meta)data. This is applied in nanoPharos as part of the provenance or bibliographical metadata. Key examples include (meta)data versioning, where the latest files include references to the previous versions as well as a record of the changes made. Curated datasets include the references, in the form of DOIs, of the original data sources and, where applicable, the DOIs of publications related to specific datasets are included in the metadata records.

3.2.4. Reusability

While FAIR data principle F2 aims to facilitate (meta)data findability, R1 corresponds to the core of the (meta)data and whether they are applicable for the intended reuse. In practice, FAIR data principle R1 refers to the scientific (meta)data. Each data point can be linked to a wide range of metadata. These include the methods, protocols, or assays used to produce them, different experiments linked to a specific material, distinct batches of a material and respective physicochemical characterisations and assays, computational descriptors, and more.

(Meta)data availability in nanoPharos is clearly defined via licensing. The default nanoPharos license is CC-BY-4.0 (Creative Commons Attribution 4.0 International), although data owners can apply their preferred license based on specific requirements, e.g., embargoed, sensitive data. The licenses are complemented with detailed provenance metadata (FAIR data principle R1.2), as described for FAIR data principles F2 and R1, which are publicly available. In this way, (re)users can identify any changes made to the datasets, when they were produced, the data owners (in nanoPharos data ownership remains with the data producers or curators for literature data), the required credit (license information), data processing methods applied, and more. All this information is based on continuous consultation with domain experts and are updated, expanded, refined as projects and materials science in general evolve (FAIR data principle R1.3).

4. Results and impact

nanoPharos provides a FAIR-compliant environment for the streamlined discovery, comparison, and reuse of NMs datasets. It offers researchers access to structured machine-actionable metadata and a rich searchable interface to enable quick hypothesis testing, aiding in the replication of experiments, and identification of knowledge gaps through data analysis. Example queries could include: "Nanosilica samples characterised by DLS and TEM," or "ZnO *in vitro* cytotoxicity studies at pH 7.4." In these cases, the searches would retrieve data from specific NMs, which contain relevant scientific metadata that are included as search terms, e.g., DLS, TEM, cytotoxicity, pH.

Similarly, nanoPharos can promote digitised R&I and R&D product development and regulatory compliance by hosting high-quality and up-to-date NMs information, with its potential increasing as more datasets are integrated. Companies can benchmark against existing materials, explore SSbD alternatives, and drive materials redesign processes. Example queries in the SSbD context might include: "Cytotoxicity of iron oxide NMs based on size and surface coating for drug delivery," or "TiO_2 NMs with ecotoxicological data."

Finally, risk assessors, regulators and policy makers can benefit from centralised, high-quality data on NMs properties and potential risks, which can be used as complementary to regulatory evaluations. nanoPharos can help with setting evidence-based safety thresholds, tracking compliance, and identifying emerging concerns, especially in cases where standardised operating procedures (SOPs) and Good Laboratory Practice (GLP) have been used and are catalogued in the dataset's metadata. Example queries could include: "NMs with human toxicity endpoints" or "Materials with environmental persistence above a threshold value (X)."

4.1. Implementation details

The current version of nanoPharos offers stakeholders a FAIR compliant database focussing on delivering high-quality ready-for-modelling NMs data. The main aim of nanoPharos is to promote the digitisation of the R&I and R&D processes of advanced and novel materials, including NMs, design and redesign, as well as for the development of digital SSbD strategies and alternative testing strategies (EC, 2024). Fig. 7 schematically presents the process by which (meta)data are published in nanoPharos. Curators access the system and upload a template containing the data they want to import. Each data point is assigned a unique internal identifier. During upload, curators link the data to a specific NM either by choosing an existing material, creating a new batch of an existing material, or by defining a new one. Defining a new material requires providing or choosing its crystallographic information, e.g., unit cell dimensions, space group etc.

At the same time, curators need to import the required metadata for transparency, reproducibility, data quality, and FAIRification purposes. Using the specialised templates provided by NanoPharos, curators define the required basic technical, descriptor specific, provenance, and bibliographic metadata. In this way, database managers and administrators can access the necessary information to ensure that data are published (or not) and to develop the machine-actionable metadata templates in nanodash. As soon as a template is submitted the system checks that the provided (meta)data meet the necessary type and format requirements and if the quality check is passed the (meta)data are imported into nanoPharos. (Meta)data are linked together using their unique identifiers to create a combined dataset. Prior to publication internal checks are performed regarding the licensing and data ownership. Based on the submission, either the entire dataset, or a summary of the metadata is published., i.e., title, summary, type, and nanodash template. The datasets are also transformed into machine-actionable JSON and XML templates.

Curators are offered the option to enrich their datasets with a series of atomistic, molecular, structural, and periodic-table descriptors, where available. These are readily available in internal libraries and can be imported to create a new version of the dataset. The enriched dataset is assigned a new identifier and published, while the metadata record is enriched with the new information included and the changes made. Similarly, if a data producer wants to update a submitted dataset, they upload it as a new version with a new assigned identifier and a record of the changes made compared to the original submission. All different dataset versions are linked under an overarching identifier for transparency and monitoring purposes. In this way, if a specific version of the dataset has been used in model development, publications, reporting, or other resources, users can access the correct version to reproduce or verify the respective results.

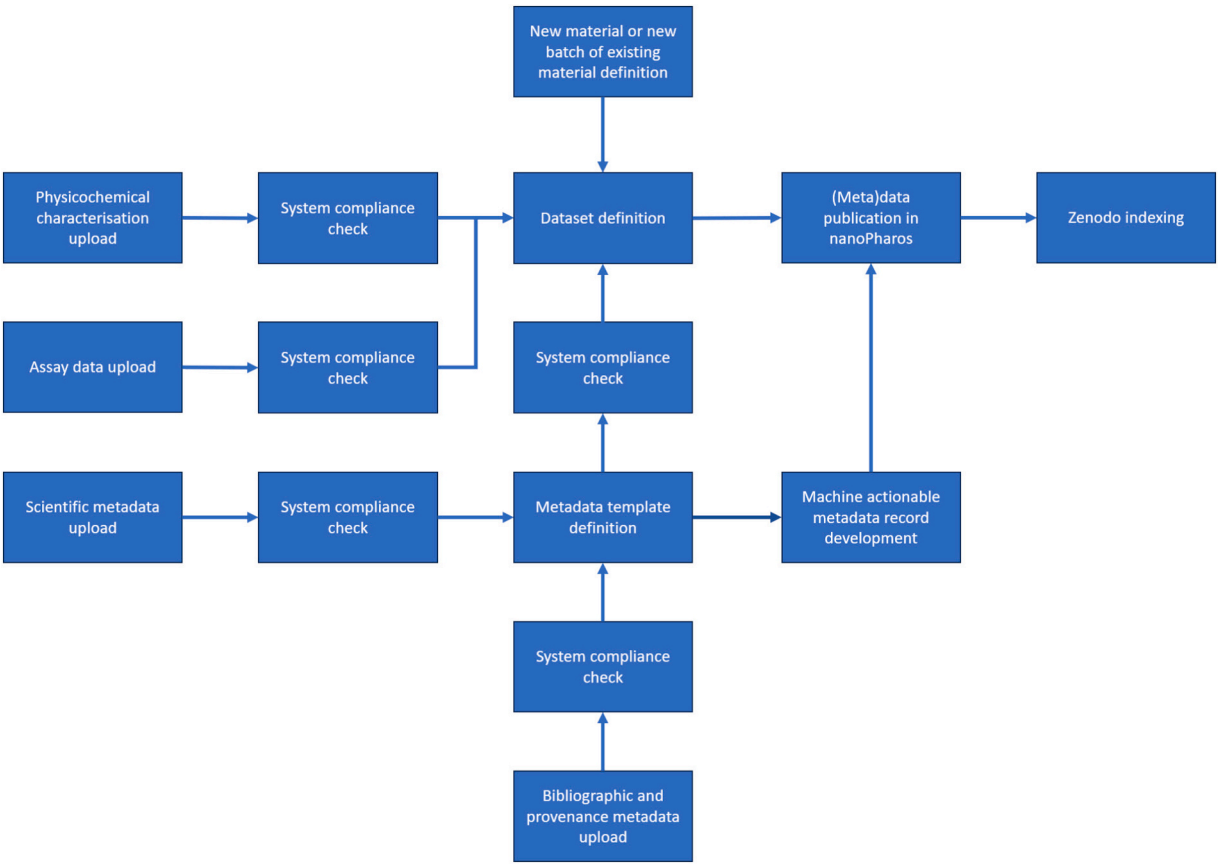

Fig. 7. Schematic workflow for (meta)data publication in nanoPharos.

4.2. FAIRness evaluation

As described in the Methodology for FAIRification section, the current version of nanoPharos is FAIR compliant and follows clearly defined processes to deliver high-quality and FAIR datasets. As projects and materials science evolves, mechanisms are put in place to maintain and expand the FAIRness of nanoPharos. These mechanisms are based on a set of quantitative and qualitative metrics. Quantitative metrics include the analysis of the metadata completeness and richness for each dataset and how well they comply with the FAIR Data Principles requirements and specific community standards. This requires continuous community and user feedback, which is achieved as part of project wide consultations, or through communication with individual users.

Furthermore, nanoPharos is being continuously evaluated, in accordance with the GO FAIR Foundation FAIR Data interpretations (GoFAIR, n.d.) and the JRC FAIR Data Guidelines (Lowenthal et al., 2025). Based on the evaluation of the current-state of nanoPharos (see Table S1 in the Supplementary Information file), out of the 41 maturity indicators nanoPharos fulfils 33 (81 %). Four (n = 4, 10 %) of the requirements, i.e., Data includes references to other data, Metadata includes references to other data, Data includes qualified references to other data, Metadata include qualified references to other data, are addressed on a case-by-case basis. This means that where applicable these conditions are met, while the JRC evaluation grid expects that it is fully implementable. One (n = 1, 2 %) has been partially implemented, i.e., Data is accessible through an access protocol that supports authentication and authorisation. This has been implemented for the human actionable (web) interface, but it is not yet supported by the nanoPharos API, although this is something that will be addressed in a subsequent iteration.

Three (n = 3, 7 %) conditions are not met, namely, Data uses knowledge representation expressed in a standardised format, Data uses machine-understandable knowledge representation, and Data uses FAIR-compliant vocabularies. For the purposes of categorisation, these results mean that the nanoPharos dataset cannot, currently, be categorised in any of the JRC categories if machine actionability using the REST API is considered. However, the web version of nanoPharos can be categorised in the *FAIR Play* (second) JRC category. The implementation of authentication and authorisation API functionality and ontologies integration are key to futher enhancing the FAIRness of NanoPharos.

On the other hand, the metadata offered by nanoPharos can be considered fully FAIR and falls under the *FAIRest of them all* category. It must be noted that some of the metadata indicators are met using 3rd party tools, i.e., nanodash, Zenodo, leading to a FAIR nanoPharos ecosystem rather than a standalone database. This is to be expected, as the FAIR Data Principles and respective interpretations, require indexing and longevity plans that rely, in practice, on separate resources. The inclusions of the machine-actionable metadata records within nanoPharos further emphasise its FAIRification.

The first version of nanoPharos (June 2020) met 10 (24 %) of the JRC maturity indicators. These were:

- Data is identified by a persistent identifier.
- Data is identified by a globally unique identifier.
- Metadata includes the identifier for the data (mainly through peer-reviewed publications).
- Metadata contains information to enable the user to get access to the data.
- Data can be accessed manually (i.e., with human intervention).
- Metadata identifier resolves to a metadata record, i.e., publication DOI.
- Data identifier resolves to a digital object.
- Metadata includes references to other metadata.

- Plurality of accurate and relevant attributes are provided to allow re-use.
- Data complies with a community standard.

Compared to the first version of NanoPharos, the current version, released in June 2024, presents a 210 % increase in the maturity indicators addressed (see Table S2 in Supplementary Information), reaching 81 % as noted above. Work is ongoing to fully meet the entirety of the maturity indicators and offer the community a fully FAIR compliant database. These results demonstrate the continuous improvement of nanoPharos and the commitment of the development team to data FAIRification.

4.3. Case studies

In parallel with the development and FAIRification work of nanoPharos itself, substantial work is taking place in capturing, digitising, and exploiting results from publicly funded projects to maximise the data available for reuse via nanoPharos. Linking of nanoPharos with other tools and platforms for automated data processing, analysis, and model development is also ongoing. Here we present key case studies to demonstrate the implementation of nanoPharos in everyday research, data digitisation, and FAIR implementation.

4.3.1. DIAGONAL instance

As part of the expansion and increase in FAIRification of nanoPharos, a dedicated instance was created to handle and accommodate the (meta) data produced from the Horizon Europe project DIAGONAL (DIAGONAL, 2021). DIAGONAL explored the implementation of the SSbD principles in advanced materials, focussing on multicomponent NMs (MCNMs) and high aspect ratio NMs (HARNs). MCNMs and HARNs present substantial research and scientific gaps, especially for in vitro experimental and modelling data (Papadiamantis et al., 2024). For this reason, DIAGONAL focussed on metal oxide MCNMs, i.e., ZnO MCNMs doped with one or more rare earth elements, Cerium-zirconium mixtures ($Ce_xZr_yO_2$) as exemplars for transition metal doping, the titanium carbide (TiC) NMs contained in Ti-6Al-4 V alloy powders, and carbon-based NMs (graphene, carbon nanotubes) and Ag nanowires as exemplar HARNs. DIAGONAL produced diverse data regarding the physicochemical characterisation, in vitro assessment, and ML model development for the SSbD and re-design of the studied materials. As these data were different, compared to those of legacy NMs imported into nanoPharos till then, and specific embargo and commercially sensitive data requirements applied, a separate instance was required, which was then mapped to the underlying nanoPharos schema.

To accommodate the experimental and computational data produced by DIAGONAL in nanoPharos, a revision of the data model was required to accommodate new data types and material information. The DIAGONAL instance was also used to expand the FAIRness of nanoPharos through the inclusion of rich bibliographical, provenance, and scientific metadata. In this way, it was possible to demonstrate the scalability and flexibility of nanoPharos to accommodate emerging materials, while enhancing its FAIRness compliance.

One key expansion of nanoPharos was the curation of rich metadata as per the FAIR Data Principles requirements (Wilkinson et al., 2016) and the respective interpretations by the GO FAIR Foundation (GoFAIR, n.d.). The new types of metadata captured included the data producers, owners, and curators per dataset. These people are identified using unique identifiers like ORCID. These can be used to contact the relevant persons in case any issues with the data are identified, a user wants to request access to embargoed data, etc. Other metadata include a short abstract of the dataset for re-users to understand the initial scope for which the data were produced, as well as a summary of the methods and protocols used to produce them, the instruments and software used for capturing or processing etc. Dataset versioning was also introduced for transparency and quality purposes to accommodate intra-project experimental evolution, data production, and corrections. All these metadata have been linked to each datapoint maximising its FAIRness potential and allowing for better reuse of specific metadata as data points during data development or meta-analysis. Relevant publications were also linked to datasets by incorporating the respective DOI into the metadata records.

Through this effort, (meta)data digitisation, storage, FAIRification, and retrieval have been streamlined for the DIAGONAL stakeholders. The expansion of the data model and the incorporation of specialised tables for the (meta)data allows researchers to filter for specific MCNM or HARN properties and retrieve relevant in vitro toxicity results in a single query. The (meta)data are then offered in a tabular format for direct import into analytical or modelling workflows. This leads to substantial savings in terms of effort and time compared to fragmented searches, even if at first the system seems more complicated. Similarly, regulators and industry partners in DIAGONAL gain a clearer overview of possible safety implications, as the database centralises NM composition, dimensionality, and risk information. These are directly linked to metadata and references and can be evaluated in terms of quality, completeness, and regulatory acceptance (Marchese Robinson et al., 2016).

In terms of FAIRness, the newly introduced attributes and advanced search functionality assists with Findability and retrieval. Persistent identifiers and rich metadata curation provide consistency and transparency for findability, interoperability, traceability and reusability purposes. Metadata richness assists with data re-use in terms of meta-analyses and in future projects that may study similar materials.

During implementation, the DIAGONAL instance was mapped to the nanoPharos model and new attributes were integrated into it. In this way, as soon as data become Open, e.g., following publication or lapse of any embargo period, the data will be available through the main nanoPharos database with a flag that they originate from the DIAGONAL project. Any new functionality introduced through the DIAGONAL instance was also inherited into the main database and is now considered as part of the nanoPharos baseline data model. This exercise also provided specific directions regarding future expansions towards in vivo data and more sophisticated synergy analyses, e.g., interactions between multiple components in MCNMs, further extending nanoPharos's applicability.

4.3.2. CompSafeNano instance

Further expansion of nanoPharos took place as part of the CompSafeNano project (Zouraris et al., 2025). The CompSafeNano project is a Research and Innovation Staff Exchange (RISE) funded project, which is heavily focussed in the integration of data-driven and physics-based models and in vitro predictive toxicology for the design of safer and sustainable NMs at the earliest stage of development (Zouraris et al., 2025). CompSafeNano relies heavily on the reuse of existing data, which are curated from literature, processed, and imported into data-driven (ML) workflows. As these data originate from different resources, i.e., publications, databases, there was increased need for richer metadata curation to accommodate the original resources and reuse purposes.

In terms of FAIRification, it was decided that further machine actionability was required to comply with the FAIR data principles, which needed to be also inherited to the base nanoPharos database. As with the DIAGONAL project, a CompSafeNano instance was developed for storing the data curated and produced within the project. Besides an expansion similar to that described in the DIAGONAL case study, distinct digital metadata records and indexing in Zenodo was introduced. These records were published as a nanopublication in nanodash (Fig. 8), and annotated using established ontologies, e.g., eNanoMapper, CheBI. The template includes the GUPRIs of the nanopublication in nanodash and the dataset in nanoPharos, its title, description, and publisher. It also includes an index of the NM ID, a reference to the protocol used to generate the data or the material using different synthesis routes, and the types of data included. The template provides

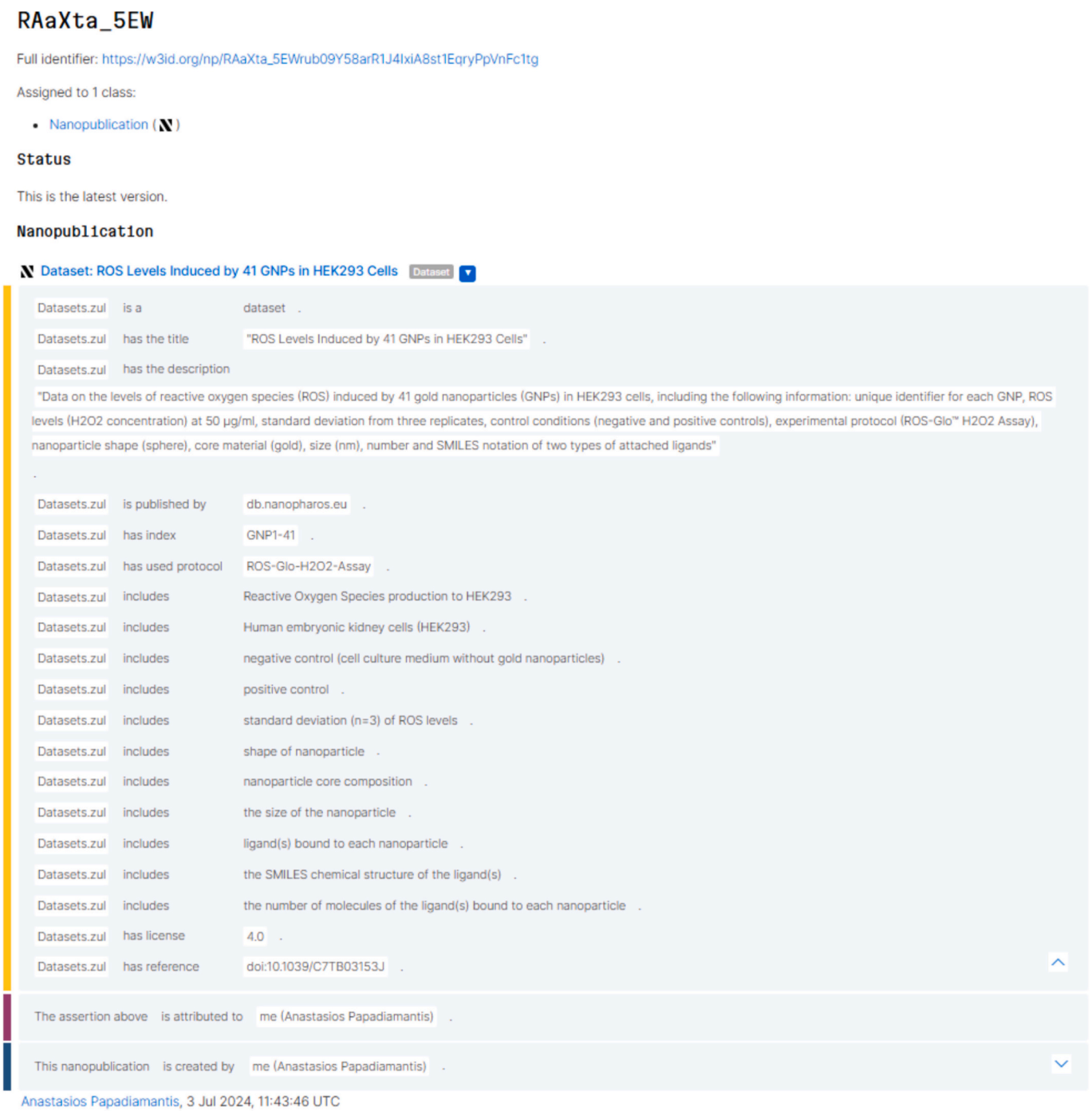

Fig. 8. nanodash template for the publication of the metadata related to the literature curated dataset titled "ROS Levels Induced by 41 GNPs in HEK293 Cells" included in nanoPharos.

information on the license of the dataset (CC-BY-4.0) and references the publication from which the data were curated. Finally, the template provides information on the person that created and published the nanopublication using their ORCID. This template is automatically made available, besides the human readable form presented in Fig. 8, as TriG, JSON-LD, N-Quads, and XML. Furthermore, a backend image presenting the use of ontological terms for annotating a user-specific template can be found in Fig. S2 of the Supplementary Information. The full backend assertion template can be found in reference (Papadiamantis, 2025) (please note that login in using an ORCID ID is required).

Following creation of the metadata record, an entry was created in Zenodo (Fig. 9) for indexing of the (meta)data. This entry includes direct references to the nanoPharos database and nanodash, the DOI assigned by Zenodo, its version in Zenodo, the Funding body, and the automatic indexing in OpenAIRE (for EU-funded research outputs). Finally, the dataset published in nanoPharos (Fig. 10) is updated with all relevant information, i.e., the dataset title and description, the resources from where the data were curated, the link to the nanodash template, and the Zenodo indexing. The literature curated data are automatically openly published in nanoPharos, while the same process will be followed for data produced within the project that may initially be closed until publication, full exploitation by the data producers, or until the defined embargo period lapses.

The implementation of these functionalities provides stakeholders

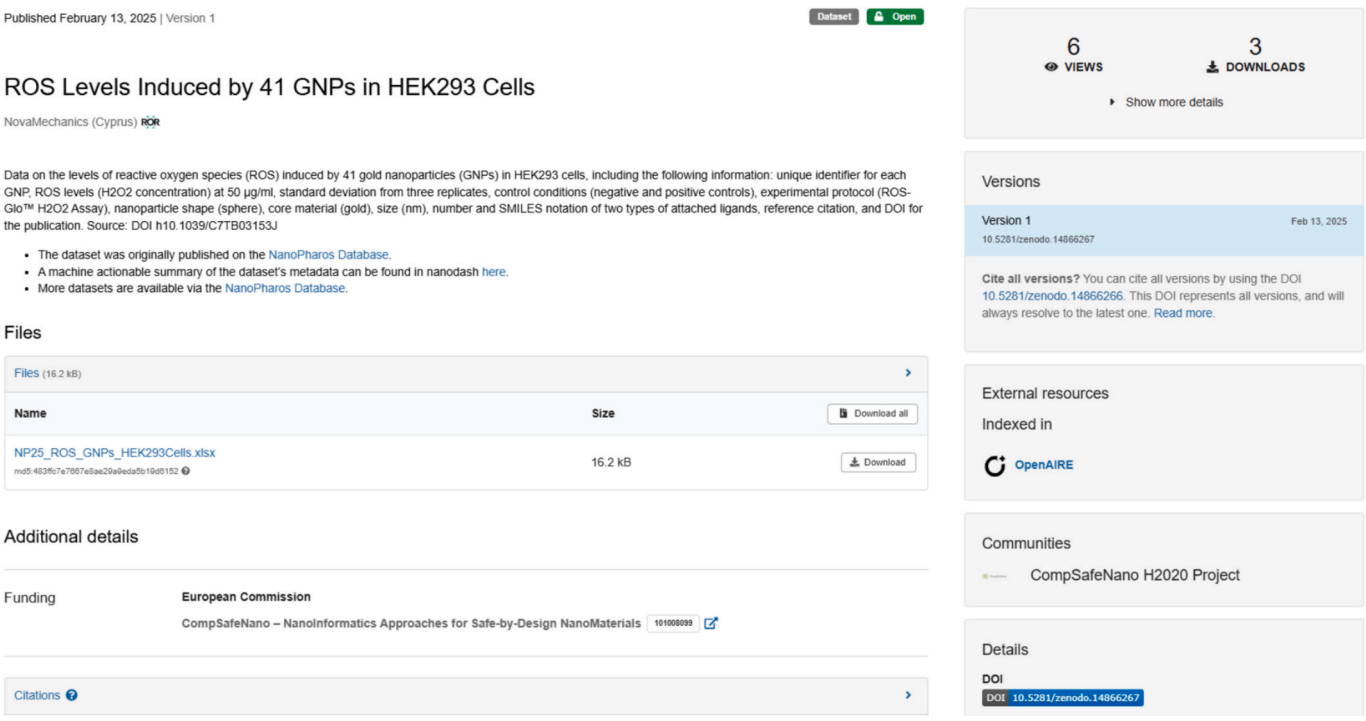

Fig. 9. Indexing of the "ROS Levels Induced by 41 GNPs in HEK293 Cells" dataset in Zenodo. The dataset received a DOI, while qualified references for the (meta) data are provided that contain their unique identifiers. Through Zenodo, the dataset is also automatically published in OpenAIRE, while information on the funding body is also provided.

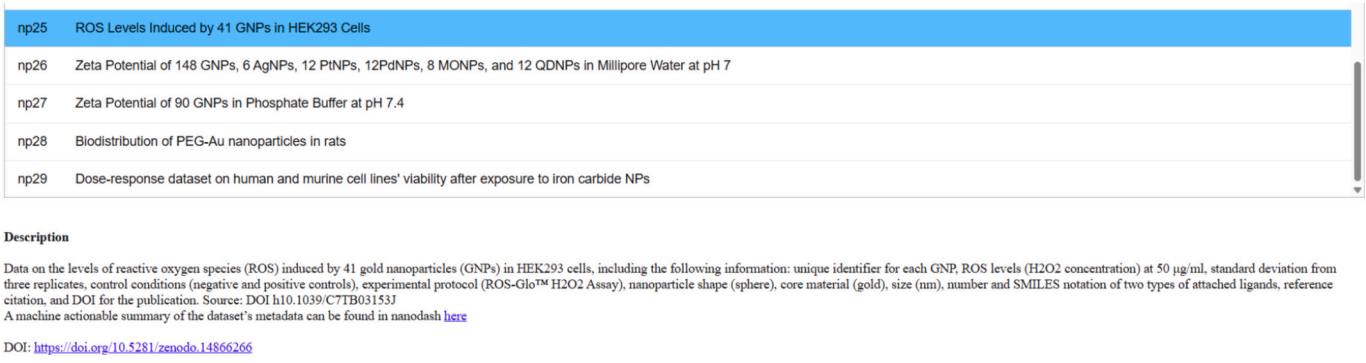

Fig. 10. Publication of a literature curated dataset in nanoPharos. The publication includes the dataset's title and description, references the original data sources, links to the nanodash metadata template, and the Zenodo indexing.

with datasets that can originate from diverse sources, but have been harmonised and cleaned, and are available from a single platform. *Re*-users can access these datasets and import them into computational workflows without any additional preprocessing on their side. The implementation of extended and machine-actionable metadata records promotes transparent and validated data sharing. Furthermore, the work performed in CompSafeNano substantially increased the nanoPharos compliance with the FAIR Data Principles. Besides the nanoPharos GUPRI, each dataset receives a persistent identifier (DOI) via Zenodo, making it globally resolvable and easier to cite. Detailed metadata, recorded both in human-readable forms and machine-actionable formats (JSON-LD, TriG, N-Quads, XML) in nanodash, boosts discoverability and facilitates automated data integration. The licensing and provenance information clarifies usage rights and streamlines future data reuse in various scientific and regulatory contexts.

4.3.3. INSIGHT – an ongoing expansion

The Horizon Europe-funded project INSIGHT ("Integrated Models for the Development and Assessment of High Impact Chemicals and Materials") is an ongoing initiative that aims to develop an integrated computational framework for assessing the health, environmental, economic, and social impacts of chemicals and materials (Serra et al., 2025). INSIGHT has established its technical foundations by creating a multi-layer system to support decision making for the SSbD of novel and advanced materials. A key element is the integration of high-quality literature data, which have been curated, structured, harmonised, and uploaded into nanoPharos, which is the reference database for the project, providing well-structured, FAIR, and ready-for-modelling datasets that contain physicochemical properties, toxicological endpoints, exposure information, and rich metadata for the INSIGHT Case Study materials, which include graphene, bio-based Synthetic Amorphous Silica (bio-SAS), and upconverting NMs.

To achieve INSIGHT's goals, the nanoPharos data model has been expanded to accommodate the data produced and harvested in the project. On the technical side, nanoPharos data are now curated, quality controlled (QCed), and processed to achieve high FAIR compliance, as

described in the Methodology for FAIRification section. The data are then being integrated into the INSIGHT framework through the respective project database instance. This process involves data cleansing, standardisation of units and nomenclature, and metadata semantic annotation using FAIR-compliant structured vocabularies and ontologies. The integration process includes the development of automated workflows that extract relevant descriptors, e.g., size, ζ-potential, morphology, from raw experimental data. These descriptors can then be linked to data-driven models to predict material behaviour and impact in organisms and the environment through the ready-for-modelling dataset format. These will be complemented with customised APIs to establish and facilitate data exchange between nanoPharos and the INSIGHT platform, i.e., the GUI of the INSIGHT framework.

5. Discussion

nanoPharos was originally developed to fill a gap in nanosafety research, which was the lack of high-quality, structured, and ready-for-modelling datasets. For this reason, a detailed data model was developed to describe a NM and its behaviour from the atomic to the macroscopic level. A NM is built starting from its unit cell characteristic and space group, complemented with atomic, molecular, and periodic table-based descriptors to eventually reach its physicochemical characterisation. These data were then linked to any available risk and hazard assessment data and published in a structured tabular form.

Scientific evolution and the continuously increasing complexity of R&I and R&D projects has substantially increased the requirements regarding data availability and exploitation, as also expressed by industrial stakeholders through the Materials 2030 Manifesto (*Materials 2030 Manifesto: Systemic Approach of Advanced Materials for Prosperity – A 2030 Perspective*, 2022) and the respective EC response (EC, 2024). The road to R&I and R&D digitisation requires the identification, retrieval, processing, and integration of interoperable data from difference sources. To achieve this, automation of the data query and retrieval processes are required, as the time and effort needed to achieve this manually may be prohibitive. For this reason, the FAIR Data Principles (Wilkinson et al., 2016) were developed to facilitate (meta)data identification and retrieval.

The technical nature of the FAIR Data Principles led to confusion regarding how they can be applied in everyday scientific and research practices. Similarly, the requirements for repositories to be FAIR resources and provide FAIR data required appropriate interpretation of the FAIR Data Principles as, in many cases, their application depends on the choices made by a specific community. To address this, high-level interpretation (Jacobsen et al., 2020; GoFAIR, n.d.) of the FAIR Data principles and community-specific guidance (Papadiamantis et al., 2020a) on their application by non-technical data producers have been published.

To be FAIR compliant, meaningfully contribute towards the digitisation of R&I, and promote the development of AI-based computational tools, data repositories, like nanoPharos, needed to adapt and expand their functionality. A key achievement in this expansion, for nanoPharos, has been the successful integration of rich and machine-actionable bibliographic, provenance, and scientific metadata. This integration took place through the expansion and adaptation of the nanoPharos underlying schema, which was that of ChemBL (*ChemBL Database Schema*, n.d.), to provide users with an understandable, functional, and user-friendly schema that accounts for complex NMs structures and advanced materials features.

This has enabled stakeholders, i.e., researchers, industry, and regulators, to locate, evaluate the applicability and interoperability, and reuse NM datasets. The process of identifying, retrieving, and combining data may seem cumbersome at first. In practice, the advanced search and output functionality in nanoPharos streamlines the process and delivers a structured, ready-for-modelling dataset. Furthermore, the creation of project-specific instances, such as those for DIAGONAL and CompSafeNano, demonstrated that nanoPharos can flexibly incorporate new material types, e.g., MCNMs, HARNs. It can also handle diverse data formats, and add specialised functionalities, e.g., versioning, expanded search features. These expansions confirm that the underlying data model is robust and scalable, and can meet emerging research needs while remaining extensible for future growth.

However, having such scalability and adaptability along with structured data required trade-offs. The decision to use a relational schema (MySQL) delivered performance and predictability for structured data. On the other hand, it also meant that semantic and hierarchical relationships, such as those existing in domain-specific ontologies like eNanoMapper (Hastings et al., 2015), are more challenging to integrate. Moreover, the curation of consistent metadata highlighted the complexities of creating consensus around the minimum information requirements. As bibliographic and provenance metadata are based on the GO FAIR Foundation interpretation of the FAIR Data principles (GoFAIR, n.d.), the requirements were defined to be common for all data contributors.

Another issue was the balance between the “gold standard” in metadata requirements, i.e., data producers need to catalogue everything, versus functional and realistic requirements that deliver an accessible and not overly complex database. Following consultation with NM and toxicology subject matter experts, customised templates were created and published for all available descriptors, including the metadata records. These consider the entirety of the data’s lifecycle and its reusability potential (Papadiamantis et al., 2020a) and were complemented with mandatory fields and validation scripts to improve data quality and transparency. This removed any ambiguity regarding the metadata required to realistically maximise data FAIRness and allow data reuse in computational and meta-analysis workflows. The established community feedback loops and the incremental rollout of new features, as demonstrated through the DIAGONAL and CompSafeNano case studies, was essential for integrating expert opinion and increasing data quality and transparency, while maintaining functionality and user satisfaction.

Despite the substantial improvement in the FAIR compliance of nanoPharos, as evidenced through the comparison of the JRC FAIR maturity indicators (Lowenthal et al., 2025) for the initial and latest nanoPharos deployment, there are several limitations remaining. Although ontology-based annotation, e.g., eNMO (Hastings et al., 2015), NPO (Thomas et al., 2011), ChEBI (de Matos et al., 2010), has been employed in nanopublications through nanodash (*nanodash*, n.d.) for the metadata records, the core nanoPharos database still lacks full semantic integration. This hinders the positioning of nanoPharos within one of the 5 FAIR maturity levels presented through the JRC FAIR Data Guidelines (Lowenthal et al., 2025) and prevents advanced mapping with external resources, as well as machine-to-machine querying, restricting the potential for automated data discovery across multiple platforms.

The current REST API (*nanoPharos API*, n.d.) supports retrieving entries by database identifiers only. It lacks advanced keyword- or ontology-driven search. This gap hinders customised data retrieval and interoperability with external tools, e.g., the Enalos Cloud Platform, which rely on programmatic, concept-based queries, and which is a requirement and thus being implemented during the ongoing INSIGHT expansion described in section 4.3.3. Another issue with the current REST API is the lack of authentication and authorisation processes. As a result, accessibility of embargoed or sensitive data is limited to manual access through the web interface limiting functionality and increasing the workload required.

Lastly, a key issue is data longevity. The nanoPharos metadata longevity is partially secured through external repositories like Zenodo (*Zenodo*, n.d.) and nanodash (*nanodash*, n.d.), for the machine-actionable publication of the datasets’ metadata, separately from the data, with reference to the GUPRI of the respective dataset embedded into the metadata. This ensures the machine and human actionable digital metadata records are available for the long-term supporting data

transparency and provenance. On the other hand, for the database and its data to remain robust over time and expand its functionality, requires continuous funding and internal organisational support. Without sustained engagement and a clear archiving strategy, curated resources may eventually become inaccessible and lost.

Overcoming these, will substantially increase the FAIRness level, leading to nanoPharos achieving one of the top 2 categories as defined in the JRC FAIR Data Guidelines (Lowenthal et al., 2025), i.e., FAIR Share or FAIRest of them all. Nevertheless, the developers of nanoPharos aim to acquire formal Trusted Repository status. This can be achieved based on CoreTrustSeal or ISO16363 standards and provides data stewards and database users with confidence regarding its long-term sustainability, curation, and data protection and security.

As can be seen from the referenced limitations, continuous database monitoring and evaluation is required for the identification of enhancement opportunities. The most significant development, currently being addressed as part of the INSIGHT project (Serra et al., 2025), involves deepening semantic integration within the core nanoPharos database and adoption of recognised ontologies to label and semantically annotate (meta)data attributes and define relationships. This would allow automated systems to resolve concepts like "nanomaterial shape" or "toxicity endpoint" across nanoPharos and retrieve data from different datasets. It would also allow the mapping of nanoPharos with external databases enhancing its data retrieval functionality. Moreover, improved machine-actionable workflows would make it easier to deposit newly generated data and facilitate real-time usage and results retrieval of modelling tools. Another enhancement regarding machine actionability is the full integration of file types like JSON and XML for all data entries, which is currently being applied to newly deposited data.

Additionally, knowledge graphs represent a promising mechanism for enabling flexible, concept-driven queries (Sikos and Philp, 2020). Instead of manually joining tables in a relational schema, knowledge graphs leverage semantic relationships to reveal indirect or previously unrecognised links among NMs, properties, and biological endpoints (Pavel et al., 2022). By translating the underlying database into a graph-based structure or hybrid model, nanoPharos could facilitate intuitive data exploration. Researchers could pose queries like, "Which classes of MCNMs share similar toxicity pathways?" and discover connections between multiple experimental conditions and publications. Such work is currently being developed and integrated within the INSIGHT project (Serra et al., 2025), where the project's SSbD framework develops a multi-layered knowledge graph consisting of data, model, and IOP graphs. In this framework, nanoPharos acts as the source of structured, harmonised, and FAIR-compliant data for exploitation.

In practice, this will be realised through nanoPharos's ongoing expansion through the INSIGHT project (see section 4.3.3). INSIGHT is building a multi-layered Scientific Knowledge Graph that combines three major elements, i.e., a data graph, a model graph, and an impact outcome pathway (IOP) graph. The data graph organises experimental results and literature-curated data, while the model graph connects predictive models such as quantitative structure–activity relationship (QSAR), physiologically based kinetic (PBK) models, and adapted AOPs. The IOP graph links the outputs of these models to practical impact

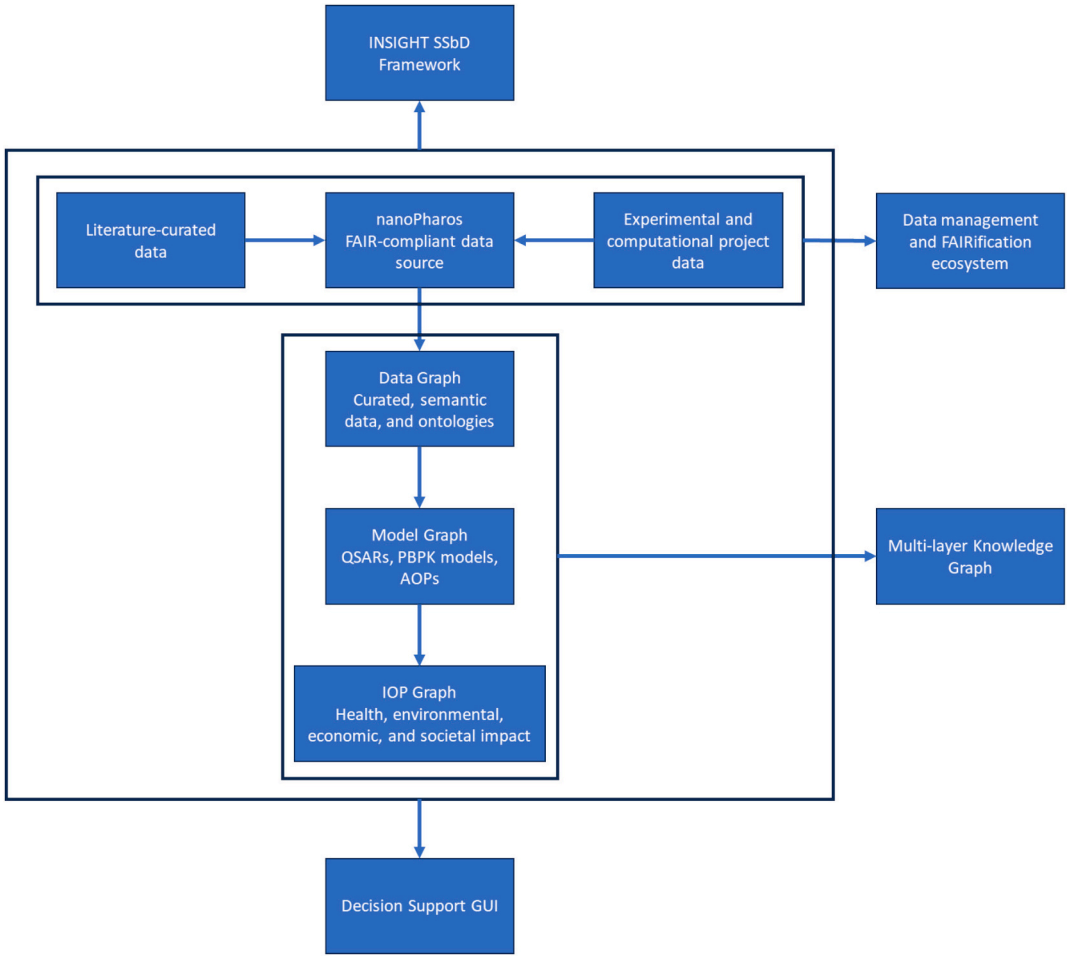

Fig. 11. Conceptual diagram showing the integration of nanoPharos within the INSIGHT SSbD framework. FAIRified NM data from nanoPharos are fed into a multi-layer Scientific Knowledge Graph structure comprising a Data Graph, a Model Graph, and an IOP Graph. These interconnected layers enable mechanistic assessment of chemicals (including NMs) health, environmental, social, and economic impacts, ultimately supporting decision-making via SSbD-aligned guidelines and tools.

indicators including regulatory endpoints, socio-economic factors, and environmental risks. nanoPharos is positioned as the framework's baseline (Fig. 11), as it will supply users with the necessary input data in a machine-readable and semantically annotated format. The ongoing mapping of the nanoPharos data to established ontologies, e.g., eNanoMapper, CheBI, and regulatory standards such as the OECD Harmonised Templates, will position nanoPharos as an active source for R&I digitisation via computational workflows and decision support tools.

Moreover, INSIGHT is extending the concept of AOPs to develop IOPs that provide a mechanistic understanding of how chemical properties lead to adverse outcomes, not only in health but also in ecological and economic contexts. The calibration and validation of the IOP models are supported by the data uploaded into nanoPharos. Application of the nanoPharos ontological and communication enhancements, as well as the expansion of its data model to fully support the INSIGHT data, will be demonstrated through the project's use cases. These test the application of graphene for energy applications, amorphous bio-based silica for automotive applications, nano-based upconverters as antimicrobial coatings, and PFAS in the aerospace industry. Integration of nanoPharos into the INSIGHT computational framework will lead to more robust and validated models, with better accuracy, and predictive ability.

Compared to other nanosafety databases and knowledge bases, e.g., NanoCommons (Maier et al., 2023), eNanoMapper (Jeliazkova et al., 2015), S^2Nano (*S2NANO : Safe and Sustainable Nanotechnology,* n.d.), Nanoinfomatics Knowledge Commons (NIKC) (Amos et al., 2021), nanoPharos's unique selling point is the provision of structured and harmonised tabular ready-for-modelling datasets. These can originate from different sources, like literature curated or project-specific datasets. While some of the databases focus on domain-specific data, e.g., toxicological, environmental, nanoPharos emphasises the maximisation of digital exploitation of available data through plug-and-play readiness for modelling and simulation workflows. Each data point in nanoPharos is harmonised in terms of, e.g., units, to minimise any preprocessing requirements. The linkage to rich metadata allows re-users to include specific metadata as data points for interoperability purposes, as demonstrated in different meta-analyses studies (Labouta et al., 2019; Bilal et al., 2019).

Similar to eNanoMapper and NanoCommons, nanoPharos aspires to implement semantic alignment internally, as well as through mapping to external databases and tools. Nevertheless, nanoPharos follows an evolutionary path which focusses on initially refining the underlying schemas and user-centric features, e.g., bulk insert, advanced search, data structuring and harmonisation, and then gradually implementing full ontology mapping. This iterative approach helps maintain a balance between immediate functionality for end users and the long-term objectives of robust machine-to-machine interoperability. For this reason, mapping to the Common European Research Information Format (CERIF) (Jeffery et al., 2014) is also planned to allow retrieval of ML and other modelling workflows that utilise nanoPharos data, and harvesting of the computationally generated data back into the nanoPharos database in a structured format thus enriching the original datasets with the generated computational data.

The lessons learned through nanoPharos's development and FAIRification can be readily adapted beyond nanosafety. Any data repository aiming to integrate rich bibliographic, provenance, and domain-specific metadata, ensure interoperability and reusability, and align with community-driven standards could adopt a similar multi-layered approach. Incrementally rolling out features like the user-centric functionality, followed by the implementation of controlled vocabularies, linking them to standardised ontologies, and finally offering knowledge graph–based queries can help with adapting to evolving and emerging digital and technical requirements without sacrificing user engagement. The modular design of nanoPharos, coupled with its reliance on scalable pipelines for data import, curation, and management defines a reproducible strategy for those who wish to modernise or adapt existing databases or build new ones that fully embrace the FAIR data principles.

6. Conclusion

nanoPharos demonstrates how a carefully designed user-centric data repository can help to accelerate digital R&I for novel and advanced materials by making data easier to access, share, and reuse for computational modelling. Its alignment with the FAIR Data Principles aims to maximise the added value of the imported data by combining and offering structured and harmonised ready-for-modelling datasets. The database uses a relational schema that contains different types of descriptors and metadata, including quality checks. This design supports a variety of users, from academia, industry, and regulation. The creation of specialised nanoPharos instances within projects like DIAGONAL, CompSafeNano and INSIGHT shows that focussing on particular research needs can expand the database to handle new types of NMs and advanced modelling workflows. Even though the database is continuously improving its FAIR compliance, there are still limitations. For example, fully incorporating ontologies and automating data retrieval with a more advanced REST API remain challenges for the future. Going forward, efforts will aim at stronger semantic integration, possibly using knowledge graphs, and more robust programmatic tools to ensure that the database is both machine-friendly and sustainable. By sharing our experiences with nanoPharos, we hope to offer a practical template for other scientific communities interested in creating high-quality, FAIR-compliant data resources that support open, reliable research.

CRediT authorship contribution statement

Anastasios G. Papadiamantis: Writing – original draft, Validation, Formal analysis, Conceptualization. **Andreas Tsoumanis:** Validation, Resources, Conceptualization. **Georgia Melagraki:** Writing – review & editing, Validation, Conceptualization. **Iseult Lynch:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Antreas Afantitis:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization.

Ethics statement

- AGP, AT, and AA are employed by NovaMechanics Ltd. a chem-, bio-, and nano-informatics company.
- AGP and IL are qualified Three-Point FAIRification (3PFF) Framework Facilitators by the GO FAIR Foundation.

Funding

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme via the DIAGONAL project (grant agreement n° 953152), the European Union's H2020 Marie Skłodowska-Curie Actions via CompSafeNano (grant agreement n° 101008099) and the European Union's Horizon Europe Programme via the INSIGHT project (grant agreement n° 101137742). This work was also funded through the UKRI Innovate UK Horizon Europe Guarantee Fund (Grant No. 10045979) as co-funding for UoB participation in the Partnership for Assessment of the Risks of Chemicals (PARC) under Grant Agreement No. 101057014.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article, including a detailed description of

the ChemBL to nanoPharos adaptation and expansion and an accompanying schematic (Fig. S1), the nanodah metadata backend Fig. S2), the computational descriptors available (Table S1) and the FAIR maturity evaluation grid (Tanle S2), can be found online at https://doi.org/10.1016/j.impact.2025.100602.

Data availability

No data was used for the research described in the article.

References

Ammar, A., Evelo, C., Willighagen, E., 2024. FAIR assessment of nanosafety data reusability with community standards. Sci. Data 11 (1), 503.

Amos, J.D., et al., 2021. The NanoInformatics knowledge commons: capturing spatial and temporal nanomaterial transformations in diverse systems. NanoImpact 23, 100331.

Amos, J.D., et al., 2024. Knowledge and instance mapping: architecture for premeditated interoperability of disparate data for materials. Sci. Data 11 (1), 173.

Ankley, G.T., et al., 2009. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29 (3), 730–741.

Baddam, P.R., Vadiyala, V.R., Thaduri, U.R., 2018. Unraveling Java's prowess and adaptable architecture in modern software development. Global Disclos. Econ. Bus. 7 (2), 97–108.

Bilal, M., et al., 2019. Bayesian network resource for meta-analysis: cellular toxicity of quantum dots. Small 15 (34), 1900510.

ChemBL Database Schema. n.d. [cited 2025 17 March]; Available from: https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest/chembl_35_schema.png.

Chetwynd, A.J., Wheeler, K.E., Lynch, I., 2019. Best practice in reporting corona studies: minimum information about nanomaterial biocorona experiments (MINBE). Nano Today 28, 100758.

DIAGONAL, 2021. Development and Scaled Implementation of Safe by Design Tools and Guidelines for Multicomponent and HARN Nanomaterials (Grant agreement no: 953152) [cited 2023 13 December]; Available from: https://www.diagonalproject.eu/.

EC, 2024. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Advanced Materials for Industrial Leadership [cited 2025 05 March]; Available from: https://research-and-innovation.ec.europa.eu/document/download/0fcf06ea-c242-44a6-b2cb-daed39584996_en?filename=com_2024_98_1_en_act_part1.pdf.

EC. Advanced Materials. n.d. [cited 2025 05 March]; Available from: https://single-market-economy.ec.europa.eu/industry/advanced-manufacturing/advanced-materials_en.

Elberskirch, L., et al., 2022. Digital research data: from analysis of existing standards to a scientific foundation for a modular metadata schema in nanosafety. Part. Fibre Toxicol. 19 (1), 1.

Ellis, L.-J.A., et al., 2018. Synthesis and characterization of Zr- and Hf-doped nano-TiO₂ as internal standards for analytical quantification of nanomaterials in complex matrices. R. Soc. Open Sci. 5 (6), 171884.

Erkimbaev, A.O., et al., 2015. A universal metadata system for the characterization of nanomaterials. Sci. Tech. Inf. Process. 42 (4), 211–222.

EU. European Green Deal. n.d. [cited 2025 05 March]; Available from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.

Exner, T.E., et al., 2023. Metadata stewardship in nanosafety research: learning from the past, preparing for an "on-the-fly" FAIR future. Front. Phys. 11.

FIP Wizard. n.d. [cited 2025 17 March]; Available from: https://fip-wizard.ds-wizard.org/wizard/.

GoFAIR. GoFAIR Foundation Interpretation of the FAIR Guiding Principles [cited 2025 06 March]; Available from: https://www.gofair.foundation/interpretation.

Hastings, J., et al., 2015. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. J. Biomed. Semant. 6 (1), 10.

Hogan, A., et al., 2021. Knowledge graphs. ACM Comput. Surv. 54 (4) p. Article 71.

Jacobsen, A., et al., 2020. FAIR principles: interpretations and implementation considerations. Data Intellig. 2 (1–2), 10–29.

Jeffery, K., et al., 2014. Research information management: the CERIF approach. Intern. J. Metadata, Semant. Ontol. 9 (1), 5–14.

Jeliazkova, N., et al., 2015. The eNanoMapper database for nanomaterial safety information. Beilstein J. Nanotechnol. 6 (1), 1609–1634.

Johnston, L.J., et al., 2020. Key challenges for evaluation of the safety of engineered nanomaterials. NanoImpact 18, 100219.

Labouta, H.I., et al., 2019. Meta-analysis of nanoparticle cytotoxicity via data-mining the literature. ACS Nano 13 (2), 1583–1594.

Lead, J.R., et al., 2018. Nanomaterials in the environment: behavior, fate, bioavailability, and effects—an updated review. Environ. Toxicol. Chem. 37 (8), 2029–2063.

Lowenthal, H., Austin, T., Da Silva, Bonino, Santos, L.O., Chiarelli, C., Cusinato, A., Ferigato, C., Friis-Christensen, A., Kemper, T., Perrotta, D., Wittwehr, C., 2025. JRC FAIR Data Guidelines. Publications Office of the European Union, Luxembourg.

Lynch, I., et al., 2020. Can an InChI for nano address the need for a simplified representation of complex nanomaterials across experimental and nanoinformatics studies? Nanomaterials 10 (12), 2493.

Lynch, I., et al., 2025. NanoPharos: Towards a Fully FAIR Database for Modelling-Ready and Computational Nanomaterials Interactions and Impacts Datasets, vol. 3. FAIR Connect, p. 2949799X251355419.

Maier, D., et al., 2023. Harmonising knowledge for safer materials via the" NanoCommons" Knowledge Base. Front. Phys. 11, 1271842.

Marchese Robinson, R.L., et al., 2016. How should the completeness and quality of curated nanomaterial data be evaluated? Nanoscale 8 (19), 9919–9943.

Martinez, D.S.T., et al., 2020. Effect of the albumin Corona on the toxicity of combined graphene oxide and cadmium to Daphnia magna and integration of the datasets into the NanoCommons Knowledge Base. Nanomaterials 10 (10), 1936.

Materials 2030 Manifesto: Systemic Approach of Advanced Materials for Prosperity – A 2030 Perspective [cited 2025 05 March]; Available from: https://www.ami2030.eu/wp-content/uploads/2022/06/advanced-materials-2030-manifesto-Published-on-7-Feb-2022.pdf.

de Matos, P., et al., 2010. ChEBI: a chemistry ontology and database. J. Chemother. 2 (1), P6.

nanodash. n.d. [cited 2025 17 March]; Available from: https://nanodash.petapico.org/.

nanoPharos API. n.d. [cited 2025 18 March]; Available from: https://db.nanopharos.eu/swagger-ui/.

nanoPharos Database. n.d. [cited 2025 06 March]; Available from: https://pharos.novamechanics.com/nanopharos.html.

NovaMechanics, 20 July 2020. NanoXtract: NanomaterialsImage Analysis Tool Powered by the Enalos Cloud Platform - A brief tutorial. Available from. http://enaloscloud.novamechanics.com/EnalosWebApps/NanoXtract/instructions.zul.

NovaMechanics Ltd, 2025. nanoPharos Data Curation Guide [cited 2025 4 November]; Available from: https://enaloscloud.novamechanics.com/diagonal/database/pages/curator/bulkInstructions.zul.

Ochoa-Zezzatti, A., et al., 2009. Improve decision support using adaptive data mining. In: 2009 International Conference on Electrical, Communications, and Computers.

OpenAIRE. n.d. [cited 2025 17 March]; Available from: https://www.openaire.eu/.

Papadiamantis, A.G., 2025. Template: Describing a dataset of ROS Production by Gold Nanoparticles in Human Embryonic Kidney (HEK293) Cells at Summary Level [cited 2025 16 October]; Available from: https://nanodash.petapico.org/publish?23&template=https://w3id.org/np/RA1XuAdO6LOtlPJgWiytJHFuK4BFHjQK5x7d9FVymzFnc&supersede=https://w3id.org/np/RAhSS6kZWfLalfUVdvQZ_b7EzykVUrfsqvz-FC42mpxqk&template-version=latest&formobj=7458878743559682851.

Papadiamantis, A.G., et al., 2020a. Metadata stewardship in nanosafety research: community-driven organisation of metadata schemas to support FAIR nanoscience data. Naomaterials 10 (10), 2033.

Papadiamantis, A.G., et al., 2020b. Predicting cytotoxicity of metal oxide nanoparticles using Isalos analytics platform. Nanomaterials 10 (10), 2017.

Papadiamantis, A.G., et al., 2021. Computational enrichment of physicochemical data for the development of a ζ-potential read-across predictive model with Isalos analytics platform. NanoImpact 22, 100308.

Papadiamantis, A.G., et al., 2024. A systematic review on the state-of-the-art and research gaps regarding inorganic and carbon-based multicomponent and high-aspect ratio nanomaterials. Comput. Struct. Biotechnol. J. 25, 211–229.

Pavel, A., et al., 2022. The potential of a data centred approach & knowledge graph data representation in chemical safety and drug design. Comput. Struct. Biotechnol. J. 20, 4837–4849.

Punz, B., et al., 2025. Instance maps as an organising concept for complex experimental workflows as demonstrated for (nano)material safety research. Beilstein J. Nanotechnol. 16, 57–77.

RFC 3986: Uniform Resource Identifier (URI). n.d. [cited 2025 17 March]; Available from: https://datatracker.ietf.org/doc/html/rfc3986.

van Rijn, J., et al., 2022. European registry of materials: global, unique identifiers for (undisclosed) nanomaterials. J. Chemother. 14 (1), 57.

S^{2}NANO : Safe & Sustainable Nanotechnology. n.d. 01 September 2020]; Available from: http://portal.s2nano.org/.

Schultes, E., 2023. The FAIR hourglass: a framework for FAIR implementation. FAIR Connect 1 (1), 13–17.

Schultes, E., et al., 2020. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. Springer International Publishing, Cham.

Serra, A., et al., 2025. INSIGHT: an integrated framework for safe and sustainable chemical and material assessment. Comput. Struct. Biotechnol. 29, 125–137.

Sikos, L.F., Philp, D., 2020. Provenance-aware knowledge representation: a survey of data models and contextualized knowledge graphs. Data Sci. Eng. 5 (3), 293–316.

Sustkova, H.P., et al., 2020. FAIR convergence matrix: optimizing the reuse of existing FAIR-related resources. Data Intellig. 2 (1–2), 158–170.

Svendsen, C., et al., 2020. Key principles and operational practices for improved nanotechnology environmental exposure assessment. Nat. Nanotechnol. 15 (9), 731–742.

The Materials Project. n.d. [cited 2025 11 March]; Available from: https://next-gen.materialsproject.org/.

Thomas, D.G., Pappu, R.V., Baker, N.A., 2011. NanoParticle ontology for cancer nanotechnology research. J. Biomed. Inform. 44 (1), 59–74.
Tsiros, P., et al., 2022. Towards an in silico integrated approach for testing and assessment of nanomaterials: from predicted indoor air concentrations to lung dose and biodistribution. Environ. Sci. Nano 9 (4), 1282–1297.
Varsou, D.-D., et al., 2020. Zeta-potential read-across model utilizing Nanodescriptors extracted via the NanoXtract image analysis tool available on the Enalos Nanoinformatics cloud platform. Small 16 (21), 1906588.
Why does the EU support research and innovation for chemicals and advanced materials? n.d. [cited 2025 05 March]; Available from: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/chemicals-and-advanced-materials_en#why-does-the-eu-support-research-and-innovation-for-chemicals-and-advanced-materials.
Wilkinson, M.D., et al., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3.
Zenodo. n.d. [cited 2025 March 17]; Available from: https://zenodo.org/.
Zouraris, D., et al., 2025. CompSafeNano project: NanoInformatics approaches for safe-by-design nanomaterials. Comput. Struct. Biotechnol. J. 29, 13–28.