



Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D4.2 FAIRification of nanoinformatics tools and models recommendations

Work Package	WP4 - Nanomaterials
Lead Author (Org)	Antreas Afantitis (NovaM) and Iseult Lynch (UoB)

Contributing Author(s) (Org)	Dimitra-Danai Varsou, Panagiotis D. Kolokathis, Andreas Tsoumanis, Georgia Melagraki (NovaM) Anastasios Papadiamantis (UoB) Thomas Exner (7P9)
Due Date	29.02.2024
Date	16.02.2024
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.10629631

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.8	02.01.2024	Antreas Afantitis and Iseult Lynch	Draft for internal review
0.9	23.01.2024	Comments from Leah McEwen - WP03 (Chemistry) and Alexander Prent - WP05 (Geochemistry)	Comments from WorldFAIR reviewers received

1.0	31.01.2024	Iseult Lynch	Content ready
-----	------------	--------------	---------------

Disclaimer

WorldFAIR has received funding from the European Commission's WIDERA coordination and support programme under the Grant Agreement no. 101058393. UK participation is funded by UKRI / Innovate UK under the Horizon Europe guarantee fund, Grant number 1831977. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

AI	Artificial Intelligence
AOP	Adverse Outcome Pathway
API	Application Programming Interface
CA	Carbonic Anhydrase (a representative protein)
CIF	Crystallographic information File
EOSC	European Open Science Cloud
EURL ECVAM	EU Reference Laboratory for alternatives to animal testing
FAIR	Findable, Accessible, Interoperable, Reusable
FAIR4RS	FAIR for Research Software (RDA Working Group)
FER	FAIR Enabling Resource
FIP	FAIR Implementation Profile
GUI	Graphical User Interface
H2020	Horizon 2020 (European Commission Framework Programme 2014-2020)
IATA	Integrated Approaches to Testing and Assessment
InChI	International Chemical Identifier (IUPAC standard)

JSON	JavaScript Object Notation
MIASE	Minimum Information About a Simulation Experiment
ML	Machine Learning
MODA	Modelling and Data (template for physics-based model documentation)
MWCNTs	Multi-walled Carbon Nanotubes
NAMs	New approach methodologies / Non-animal testing methods
NInChI	InChI extension for Nanomaterials
NMs	Nanomaterials
OECD	Organisation for Economic Cooperation and Development
PBPK	Physiologically based pharmacokinetic (models)
QMRF	QSAR Model Report Form
QPAR	Quantitative Structure-Property Relationship (model)
QSAR	Quantitative Structure-Activity Relationship (model)
RDA	Research Data Alliance
REST	Representational State Transfer
SbD	Safe by Design (now also called SSbD - Safe & Sustainable by Design)
SDF	Structure-Data File (chemical format)
SME	Small and Medium Enterprises
SMILES	Simplified Molecular-Input Line-Entry System
SOP	Standard Operating Procedure
TEM	Transmission Electron Microscopy
VO	Virtual Organisation

Executive summary

Nanomaterials, with properties of both chemicals and particles, offer exciting opportunities in a range of industrial and consumer applications, from sensing and diagnostics to precision medicine and agriculture. Paradoxically, the properties that make them advantageous for applications, including their small size and large surface area, are also the source of concerns regarding potential negative impacts arising from their uptake by, and interactions with, humans and the environment. Given the enormous diversity of nanomaterials compositions, it is not possible to individually test them using the current time, cost and animal-intensive regulatory testing approaches, driving an urgent need for alternative *in silico* approaches to predict nanomaterials safety (nanoinformatics).

The nanomaterials safety community has been actively developing a range of modelling approaches, spanning from physics-based models to data-driven approaches including machine learning models. As these models utilise and generate extensive datasets, there is a requirement for good practice in data documentation to support model development. Additionally, the models and associated software need to be FAIR (findable, accessible, interoperable and re-usable). While there is much in common with the FAIR needs for software in chemoinformatics, there are some unique aspects to nanomaterials software (nanoinformatics) that require domain-specific tailoring.

This WorldFAIR Deliverable report, which is *targeted towards nanoinformatics model developers*, presents a set of recommendations and prototypes for FAIRification of nanoinformatics tools and models. This deliverable is a stand-alone document primarily focused on FAIRification of nanoinformatics tools and software, also addressing FAIRification of the underpinning (and resulting) datasets. Organisation of the datasets into ready-for-modelling formats, for example via NanoPharos, and use of KNIME nodes to integrate the datasets directly into the modelling software, and the resulting predictions and validation statistics back into the database for further re-use are also emphasised.

This report provides an analysis of the direction FAIRification of nanoinformatics software could/should take. The report provides examples of the approaches and best practice that have emerged from Horizon 2020-funded (H2020) nanosafety-specific projects including [NanoCommons](https://cordis.europa.eu/project/id/731032)¹, [NanoSolveIT](https://cordis.europa.eu/project/id/814572)², [RiskGONE](https://cordis.europa.eu/project/id/814425)³, and [CompSafeNano](https://cordis.europa.eu/project/id/101008099)⁴ to support FAIR software. Approaches and best practice examples include the documentation of models and software via existing and emerging metadata standards, establishment of a registry of nanoinformatics models, deployment of predictive models as web applications or application programming interfaces and a

¹ H2020 research infrastructure project, NanoCommons: <https://cordis.europa.eu/project/id/731032>

² H2020 R&I nanoinformatics project, NanoSolveIT: <https://cordis.europa.eu/project/id/814572>

³ H2020 R&I nanomaterials risk governance project, RiskGONE: <https://cordis.europa.eu/project/id/814425>

⁴ H2020 Marie Curie RISE researcher exchange project, CompSafeNano: <https://cordis.europa.eu/project/id/101008099>



DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

demonstration of model interoperability and enhanced re-usability via containerisation and deployment via a cloud platform. Recommendations for next steps are provided to drive progress.



Table of contents

Executive summary	5
1. Introduction	9
1.1. Emerging principles for FAIR research software	11
1.1.1. FAIR for Research Software (FAIR4RS)	11
1.1.2. Minimum Information About a Simulation Experiment (MIASE) Guidelines	13
1.2. FAIR Principles in Nanoinformatics	15
1.3. Consideration of the barriers and drivers to FAIRification of data and software	17
2. Community standards for documentation of in silico approaches for (nano)safety assessment	20
2.1. QMRF templates as a means of documenting data-driven in silico models	20
2.2. MODA templates as a means of documenting physics-based materials models	22
2.3. Role of MODA in standardising modelling and simulation for nanomaterials	26
2.4. Role of QMRF/MODA in increasing the FAIRness of nanoinformatics models	29
3. Recommendations and tools for FAIR nanoinformatics models	32
3.1. Fully documented models / FAIR models	33
3.2. Rich metadata for nanoinformatics models	37
3.3. Registry of nanoinformatics models	38
3.4. Curation of datasets for data driven models	40
3.5. Automated integration of datasets into modelling tools via KNIME Nodes	41
3.6. Deployment of nanoinformatics models as FAIR software	42
3.6.1. Deployment as web applications	42
3.6.2. Deployment via APIs	44
3.7. Automated integration of model outputs back into nanoinformatics databases	47
3.8. Sustainability of models and software via containerisation as microservices	48
3.8.1. Docker containers for model deployment	48
3.8.2. Reproducibility and OpenAPI specification	49
3.8.3. Example: Docker in practice	49
3.9. Maximising model utilisation through provision of training materials	49
4. Examples of best practice from the nanosafety community	50
4.1. NanoCommons KNIME nodes	50
4.2. NanoCommons User Handbook	53
4.3. Deployment of models / software to EOSC	57
4.3.1. Approach to deployment	57

4.3.1.1. Cluster setup	57
4.3.1.2. Kubernetes environment	58
4.4. The NanoSolveIT cloud platform	60
4.5. Integrated workflow - from model development to FAIR software	63
5. Conclusions and recommendations	66
6. Bibliography	69
Appendix 1: NanoPharos FAIR Implementation Profile	73

1. Introduction

Nanomaterials, defined as particles with at least one dimension in the nanoscale (between 1 and 100 nm for regulatory purposes in the EU) have the properties of both chemicals and particles, offering exciting opportunities in a range of industrial and consumer applications, from sensing and diagnostics to precision medicine and agriculture. Paradoxically, the properties that make them advantageous for applications, including their small size and large surface area, are also the source of concerns regarding potential negative impacts arising from their uptake by, and interactions with, humans and the environment. Given the enormous diversity of nanomaterials compositions, it is not possible to individually test them using the current time, cost and animal-intensive regulatory testing approaches, driving an urgent need for alternative *in silico* approaches to predict nanomaterials safety (nanoinformatics).

The need to reduce animal testing, as laid out in EU Directive 2010/63/EU which introduced the terms refinement, replacement and reduction (the '3Rs'),⁵ has driven a growing awareness of the need for new approach methodologies (NAMs) or non-animal testing methods to support the risk assessment of chemicals including nanomaterials. A 2016 scientific workshop organised by the European Chemicals Agency (ECHA) on NAMs defined them as a broad category of methods that includes *in silico* approaches, *in chemico* and *in vitro* assays, providing information on the impacts of exposure to chemicals in the context of hazard assessment⁶. They also include a variety of new testing tools, such as "high-throughput screening" and "high-content methods" e.g., genomics, proteomics, metabolomics; as well as some "conventional" methods that aim to improve understanding of toxic effects, either through improving toxicokinetic or toxicodynamic knowledge for substances. The workshop recognised a need for standardisation of NAMs, including reporting templates, as well as a better understanding of their relevance through thorough analysis of their performance and definition of their applicability¹⁴.

The nanomaterials safety community has been actively developing a range of modelling approaches, spanning from physics-based models to data-driven approaches including machine learning models to support the transition away from animal testing towards *in silico* approaches. As these models utilise and generate extensive datasets, there is a requirement for good practice in data documentation to support model development. Additionally, the models and associated software need to be FAIR (findable, accessible, interoperable and re-usable). While there is much in common with the FAIR needs for software in chemoinformatics, there are some unique aspects to nanomaterials.

⁵ The 3Rs directive: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0063>

⁶ European Chemicals Agency, New Approach Methodologies in Regulatory Science. Proceedings of a scientific workshop, Helsinki, 19–20 April 2016.
https://echa.europa.eu/documents/10162/21838212/scientific_ws_proceedings_en.pdf/a2087434-0407-4705-9057-95d9c2c2cc57

Physics-based models are being developed to predict nanomaterials properties such as their surface charge (presented as zeta potential) or their binding affinities for various proteins and other biomolecules. Data-driven approaches based on read-across from data-rich materials to data poorer ones, include quantitative structure–activity relationship (QSAR) models (which for nanoscale materials are also called quantitative property-activity relationships or QPAR) as well as machine learning and deep learning models that seek to identify patterns in the data allowing correlation between nanomaterials properties and toxicity outcomes, for example. These nanoinformatics or *in silico* models generate extensive data and act as a major re-user of experimental data. There are legitimate concerns that the quality of predictions from *in silico* models depends critically on the quality of the input data, generally coming from experimental assays. Thus, good practice in data documentation, including application of the FAIR principles (findable, accessible, interoperable and re-usable) to the datasets utilised in model development is essential. FAIR nanomaterials data has been the topic of a previous WorldFAIR deliverable (D4.1; Lynch et al, 2023), while the current deliverable report covers the requirement for extension of the FAIR principles to cover software. While there is much in common with the FAIR needs for software in other fields, including chemoinformatics, there are some unique aspects to nanomaterials models and software (nanoinformatics) that require domain-specific tailoring. For example, nanoinformatics approaches may involve the integration of physics-based and data-driven software approaches that require specific validation and documentation to fulfil regulatory requirements.

Specifically in light of the need to increase regulatory acceptance of *in silico* NAM, the need to define what a model is, and to clarify the causal relationships from nanoinformatics models, were key discussion themes at the 2022 Beilstein Workshop⁷ on nanoinformatics. There was broad consensus that we need to move beyond black-box approaches and be able to demonstrate causality in predictions, which requires domain knowledge to demonstrate. A key recommendation from this discussion was to move towards multi-objective optimization whereby more than one objective function should be optimised simultaneously. In a materials context, this can be for example, optimisation of both Young’s modulus and density of a material (Gopakumar et al., 2018), while in a nanosafety context this could involve optimisation of both safety and functionality of a material.

The need for a chemical/particle-based description of the outcome was highlighted, with a suggestion that we consider quantitative *substance-activity* relationships rather than quantitative *structure-activity* relationships. So-called “explainable AI” is very much an emerging field, but computer scientists have some interesting ideas about how to make deep learning more transparent, and thus fixable and accountable, e.g., Rawashdeh (2023) who states that “There are different models for how to do this, but we essentially need a way to figure out which inputs are causing what. It is very much an unsolved problem right now”. Enhancing the confidence of

⁷ <https://www.beilstein-institut.de/en/symposia/archive/nanotechnology/nanoinformatics-2022/>

researchers, industrial users and regulators in nanoinformatics is essential also in light of the EU's draft regulations on Artificial Intelligence (AI) systems⁸. Application and extension of the emerging principles for FAIR research software is a fundamental step towards achieving this acceptance and widespread uptake of nanoinformatics models.

This document is targeted towards nanoinformatics model, tool and software developers given its focus on FAIRification of nanoinformatics software. It will also be of broad interest to data infrastructure developers, and to data curators, given that FAIR data is a prerequisite for FAIR research software.

The rest of the document is structured as follows: first we present background on emerging principles for FAIR research software and FAIR nanoinformatics software, followed by a set of recommendations regarding the key steps in the FAIRification of models, organised according to the lifecycle of model / software development and validation, and then examples of approaches to increase the FAIRness of nanoinformatics models in practice, as applied in other EU-funded research projects addressing nanosafety data and software. Finally, we provide a summary of the key messages and recommendations to the nanoinformatics community on how to implement FAIR nanoinformatics models.

1.1. Emerging principles for FAIR research software

1.1.1. FAIR for Research Software (FAIR4RS)

The FAIR for Research Software (FAIR4RS) Working Group of the Research Data Alliance (RDA) has adapted the FAIR Guiding Principles to create the FAIR Principles for Research Software (FAIR4RS Principles), as presented in Barker et al. (2022). The contents and context of the FAIR4RS Principles are summarised in Box 1 to provide the basis for discussion of their adoption and adaptation within the WorldFAIR Nanomaterials Work Package (WP04). Research software is defined by the RDA FAIR4RS WG as including, “source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose. Software components (e.g., operating systems, libraries, dependencies, packages, scripts, etc.) that are used for research but were not created during or with a clear research intent should be considered software in research and not Research Software. This differentiation may vary between disciplines”.

⁸ EU Draft Regulation on Artificial Intelligence, 2023.

<https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>

Box 1: Summary of the RDA FAIR for Research Software Principles (from Barker et al., 2022)

F: Software, and its associated metadata, is easy for both humans and machines to find

F1. Software is assigned a globally unique and persistent identifier.

F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.

F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

A: Software, and its metadata, is retrievable via standardised protocols

A1. Software is retrievable by its identifier using a standardised communications protocol.

A1.1. The protocol is open, free, and universally implementable.

A1.2. The protocol allows for an authentication and authorization procedure, where necessary.

A2. Metadata are accessible, even when the software is no longer available.

I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software)

R1. Software is described with a plurality of accurate and relevant attributes.

R1.1. Software is given a clear and accessible licence.

R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

R3. Software meets domain-relevant community standards.

Within the WorldFAIR WP04 context, the focus is on models and software for nanosafety assessment, which includes both physics-based and data driven models as defined by the European Materials Modelling Council and the European Commission⁹. There is also a need for democratisation of access to the models through provision of a graphical user interface (GUI) such that users with limited coding or computer programming skills are able to apply and run the models. This is especially important in that the intended end-users of many of the models and/or software tools are often small and medium enterprises (SMEs) producing and using nanomaterials who may not be able to employ dedicated programmers, and regulators who need to understand the physical basis of the models and the grouping principle or hypothesis underpinning the model in order to understand whether the data produced are valid for regulatory decision-making¹⁰. A key aspect of gaining regulatory acceptance of *in silico* models is documentation of the model principles and the underpinning datasets; for Quantitative Structure-Activity Relationship (QSAR) models for regulatory use, this is via the QSAR Model Report Form (QMRF), for example: see further details in section 2 below.

1.1.2. Minimum Information About a Simulation Experiment (MIASE) Guidelines

One of the main requirements in scientific research is reproducibility of experimental work. This has led to an extensive debate on the amount of information needed to be reported on for an experiment to be reproducible. As a result, the establishment of minimum information guidelines has proven valuable for promoting reproducible science in several fields. One of the first attempts was the Minimum Information Required in the Annotation of Models (MIRIAM) guidelines promoting the exchange and reuse of biochemical models (Le Novère et al., 2005). However, as reported by Waltemath et al. (2011), MIRIAM does not provide sufficient information for efficient reuse in a computational setting. To address this gap, further guidelines were devised to describe the minimal set of information that must be provided to allow the full reproducibility of a simulation experiment (Waltemath et al., 2011). These guidelines, *Minimum information about a simulation experiment (MIASE)*, include the list of models to use and their modifications, all the simulation procedures to apply and in which order, the processing of the raw numerical results, and the description of the final output. These guidelines allow for the reproduction of any simulation experiment, and are summarised as:

- I. All models used in the experiment must be identified, accessible, and fully described.

⁹ European Commission, Directorate-General for Research and Innovation, Baas, A., *What makes a material function? – Let me compute the ways – Modelling in H2020 LEIT-NMBP programme materials and nanotechnology projects – Sixth version*, Baas, A.(editor), Publications Office of the European Union, 2017, <https://data.europa.eu/doi/10.2777/417118>

¹⁰ We note that there is a dedicated programme for validation of alternative test methods (including *in silico* models) for regulatory purposes via the EU Reference Laboratory for alternatives to animal testing (EURL ECVAM), but that as yet no nanomaterials models have been validated for use, although they can be used as part of “weight of evidence”.

- a. The description of the simulation experiment must be provided together with the models necessary for the experiment, or with a precise and unambiguous way of accessing those models.
 - b. The models required for the simulations must be provided with all governing equations, parameter values, and necessary conditions (initial state and/or boundary conditions).
 - c. If a model is not encoded in a standard format, then the model code must be made available to the user. If a model is not encoded in an open format or code, its full description must be provided, sufficient to re-implement it.
 - d. Any modification of a model (pre-processing) required before the execution of a step of the simulation experiment must be described.
- II. A precise description of the simulation steps and other procedures used by the experiment must be provided.
 - a. All simulation steps must be clearly described, including the simulation algorithms to be used, the models on which to apply each simulation, the order of the simulation steps, and the data processing to be done between the simulation steps.
 - b. All information needed for the correct implementation of the necessary simulation steps must be included through precise descriptions or references to unambiguous information sources.
 - c. If a simulation step is performed using a computer program for which source code is not available, all the information needed to reproduce the simulation, and not just repeat it, must be provided, including the algorithms used by the original software and any information necessary to implement them, such as the discretization and integration methods.
 - d. If it is known that a simulation step will produce different results when performed in a different simulation environment or on a different computational platform, an explanation must be given of how the model has to be run with the specified environment/platform in order to achieve the purpose of the experiment.
- III. All information necessary to obtain the desired numerical results must be provided.
 - a. All post-processing steps applied on the raw numerical results of simulation steps in order to generate the final results have to be described in detail. That includes the identification of data to process, the order in which changes were applied, and also the nature of changes.

- b. If the expected insights depend on the relation between different results, such as a plot of one against another, the results to be compared have to be specified (Waltemath et al., 2011).

1.2. FAIR Principles in Nanoinformatics

Nanoinformatics is defined as the application of informatics to nanotechnology. Dedicated application of the FAIR principles is expected to enhance the nanoinformatics tools and software directly, as well as critically, optimise the underlying datasets that are integral to the functionality of data-driven models (Afantitis *et al.*, 2020), enabling maximal efficacy and utility by computational tools in the application of informatics to nanotechnology.

The performance and accuracy of data-driven nanoinformatics models, such as machine learning algorithms and QSAR models, are intrinsically linked to the quality, organisation, and accessibility of their underlying datasets. These models rely heavily on meticulous data curation (i.e., data cleaning, structuring, and standardisation) of nanomaterials datasets for optimal functioning.

Platforms such as [NanoPharos](#)¹¹, which have been developed intentionally to fill a gap in terms of access to ready-for-modelling datasets, play a pivotal role by providing structured frameworks conducive to the management of nanomaterials data. They facilitate the transformation of datasets into forms that are not only aligned with the FAIR principles but are also primed for efficient and precise analytical processing via computational modelling. Further enhancing this process is the integration of these organised datasets into modelling tools, a task that is easily managed by existing widely-used platforms like [KNIME](#)¹² (KoNstanz Information MinEr), an open-source platform for data analytics. KNIME leverages nodes - its fundamental processing units - to automate data workflow processes. This automation is of particular benefit in nanoinformatics, enabling the direct integration of datasets from repositories like NanoPharos into various modelling software and allowing for a fluid transition from the phase of data curation to practical application in modelling. KNIME is fully open source¹³ and thus provides an excellent basis for enhancing the FAIRification of models - indeed, WorldFAIR Deliverable report D4.1¹⁴ (Lynch et al., 2023) provided an analysis of the use of KNIME as a means to make datasets FAIR, including automating the annotation of datasets with canonical identifiers such as the IUPAC standard [InChI](#) for chemicals¹⁵ along with the proposed NanoInChI for nanomaterials (Lynch et al., 2020). We note

¹¹ NanoPharos database of modelling-ready datasets: <https://db.nanopharos.eu/Queries/Datasets.zul>

¹² KNIME, <https://www.knime.com/>

¹³ KNIME is written in Java and based on Eclipse. It is an open source multi-language software development environment with an integrated development environment and an extensible plug-in system. KNIME Analytics Platform is released under an Open Source GPLv3 licence that allows others to use the node API to add proprietary extensions.

¹⁴ <http://doi.org/10.5281/zenodo.7887340>

¹⁵ InChI or International Chemicals Identifier, a representation for chemical structures; <https://www.inchi-trust.org/>

that an InChI string itself does not resolve to a persistent location but as the notation is canonical the same string will always be generated for a given compound (within a very small error of collisions). KNIME can also support enrichment of datasets with ontology terms and community- or user-defined metadata, provide extensive metadata about workflows (exportable workflow summaries), and also includes the KNIME hub where workflows can be made publicly and persistently available by and for any user, ensuring transparency, reproducibility and provenance.

The utility of KNIME-linked nanoinformatics models is further realised in the post-processing phase, where the generated predictions and validation statistics are critical in evaluating model precision and reliability. An automated mechanism for channelling these results back into databases such as NanoPharos not only augments the data repository but also provides invaluable insights for subsequent research initiatives. This process involves the augmentation of the database with new research findings, predictive accuracies, and any ancillary metadata generated during the modelling process, thereby enhancing the datasets' reusability and overall value.

A recent publication, including co-authors from WorldFAIR WP04, outlined the ambition of the GO FAIR Foundation-linked AdvancedNano Implementation Network to support community wide-adoption of the FAIR principles, as depicted in Figure 1 (Dumit et al., 2023). The three pillars of the action plan are: **1) Definition and set-up**, where domain-specific descriptions of FAIR implementation principles are developed that specify data and software that are technically fit for the nanosafety community and its down-stream research and innovation fields; **2) Implementation**, whereby new solutions, methods and tools are employed and endorsed as *FAIR Enabling Resources (FERs)* within and beyond the nanosafety community e.g., through the development of case studies for monitoring, refinement and improvement of solutions; and finally, **3) Roll-out and operation**, whereby principles, solutions and tools are embedded into practice through widespread communication, training and education initiatives, such as [Elixir TeSS](#).

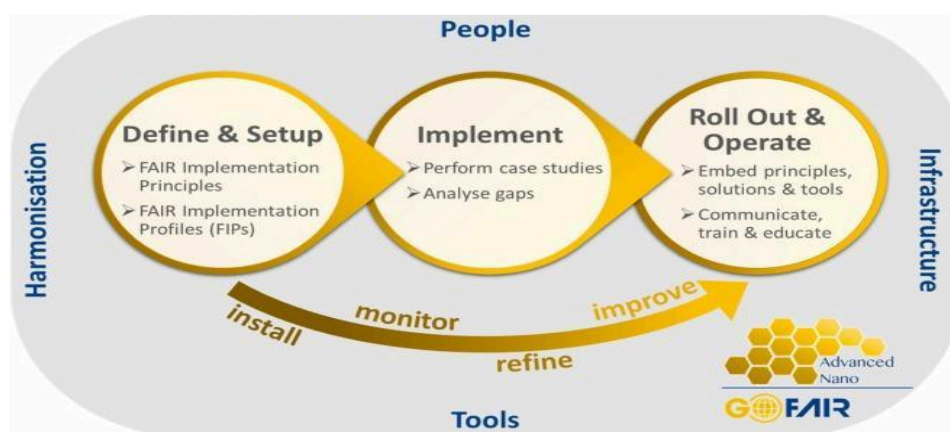


Figure 1. The AdvancedNano IN action plan foresees three main action categories at the centre of the identified FAIRification challenges, *i.e.* people's awareness, harmonisation, tools, and infrastructure, and involves monitoring strategies enabling refinement and further improvement of the FAIRification process. From Dumit et al., 2023.

1.3. Consideration of the barriers and drivers to FAIRification of data and software

Several reports from nanosafety projects have discussed the barriers to FAIRification of data, including a report from the Horizon 2020 project [EC4SafeNano](#) in 2019, which confirmed that “few, or none, of the issues related to data sharing are nanomaterials specific, and thus there is enormous scope for learning from, and leveraging of best practice in data management and data sharing from other contiguous areas.”¹⁶ From the literature analysis performed for the EC4SafeNano report, six categories of barriers were identified: technical, motivational, economic, political, legal, and ethical barriers (van Panhuis et al., 2014). Interestingly, the nanoinformatics community have not had similar discussions or publications on barriers to model sharing, likely as model / software developers generally want user to utilise their software, and due to the general culture of open source code whereby software is released under a licence in which the copyright holder grants users the rights to use, study, change, and distribute the software and its source code to anyone and for any purpose. That said, in practice, collaboration between nanoinformatics modellers has been through utilising common datasets on which to develop and compare models rather than sharing code for independent application on other datasets, at least prior to publication of the model and its outcomes in the literature.

The R of FAIR, i.e., re-use of data, has been the most controversial to date. As discussed in a recent paper on metadata for nanoinformatics from WorldFAIR WP04 (Exner et al., 2023), reuse of data is a complicated concept and the exact meaning of “reuse of research data” (and what is needed to be able to do so) varies between disciplines and individuals with no common standard applied. Efforts to evaluate criteria that distinguish reuse from other related actions and to find a definition that makes reuse measurable and understandable to a broader audience have either:

1. Distinguished the three variables: research question, research data and research method to enable a clear differentiation between reuse and related concepts. According to this view, the reuse of data enforces the usage of the same data with a different method and a different research question in mind (Schöch, 2017)¹⁷; or
2. Focussed on identifying the characteristics of reuse by examining the etymology of the term and analysis of the current discourse, leading to a range of reuse scenarios that show the complexity of today’s data-driven research landscape (Figure 2). This approach suggests that there is *no reason to distinguish use and reuse*. In this re-conception, (re)use is defined as the use of any research resource (including software in the context of the current

¹⁶ EC4SafeNano, 2019. Report on mechanisms to overcome barriers to collect and make available the data. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ca75a401&appId=PP_GMS. Page 3.

¹⁷ Schöch C. Wiederholende forschung in den digitalen geisteswissenschaften (2017). Available at: <https://christofs.github.io/wiederholende-forschung-dhd/> (Accessed August 4, 2023).

deliverable) regardless of when it is used, the purpose, the characteristics of the data and its user. This reflects the fact that the research landscape is no longer linear, but rather a complex and dynamic landscape where “character of data”, “the user”, “purpose” and “time” are no longer solid pillars, and thus use and re-use are interchangeable (van de Sandt et al., 2019).

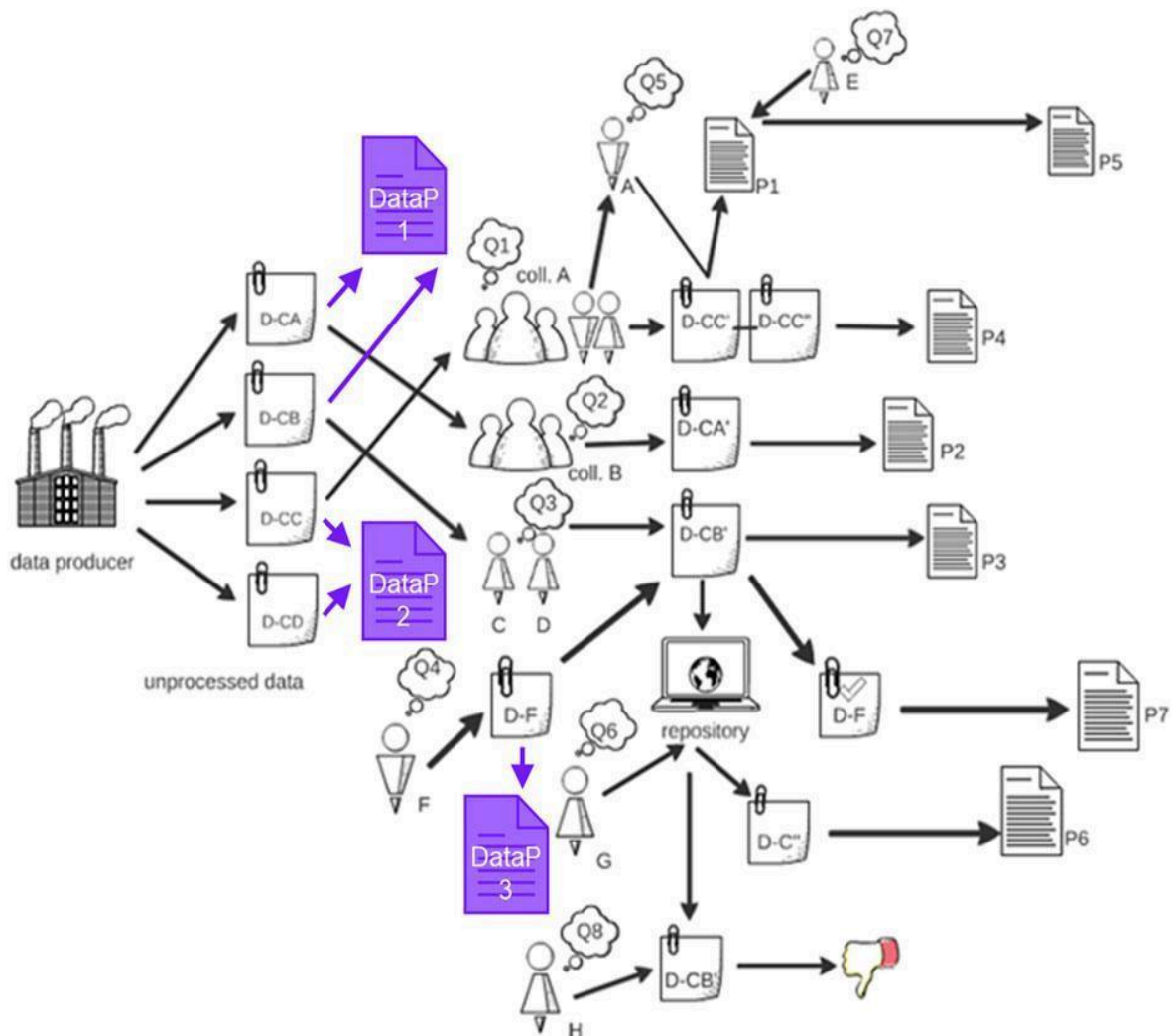


Figure 2: Schematic illustration of the concept of reusable research and the complexity of (re-)use scenarios: D are various datasets (D-), which are used alone and/or integrated with other datasets and used by different individuals and collaboration teams to address a range of questions (Q) leading to a set of publications (P). Importantly, characteristics like “character of data,” “user,” “purpose” and “time” are no longer solid pillars, and the process of data use and re-use is no longer linear as these now occur in parallel and progress at different rates. In the context of models and software,

DataP could be replaced by a software paper¹⁸ as a means to document and share software and enable citability, provide a DOI and support FAIR software. Reproduced from Exner et al., 2023.

In this deliverable, we adopt the non-linear approach as described in the second case and consider the data management and FAIRification requirements to be broadly the same irrespective of whether researchers are undertaking FAIRification to support first analysis of their own data or to support (re)use of data or models by others. We acknowledge that the incentives and need for FAIRification are seen as being lower for data providers than for data re-users, but effective data (and model) sharing can only work with the commitment of those producing the data (experimentalists) and software (developers). Reuse of data is facilitated through the development and widespread application of domain-specific community standards for rich metadata. These provide sufficient context to how and why the data were generated to allow other researchers to determine if the data is suitable to also address their specific research question. For example, Elberskirch et al., (2022) proposed a specification for the necessary minimum information to be provided along with experimental results on effects of nanomaterials in the biological context that is divided into six modules: general information, material information, biological model information, exposure information, endpoint read out information and analysis and statistics. In the case of reuse of nanosafety data for regulatory purposes, e.g., as part of the weight of evidence for whether a nanomaterial is toxic or not under specific use conditions, much of the required metadata relates to the specifics of the Standard Operating Procedure used to generate the data and whether it was generated in accordance with an approved OECD Test Guideline study for example, as well as considerations around data quality and completeness (see, for example, Drobne et al., 2023).

Given that models and software are designed to be used by their community, we anticipate that there will be a significant incentive for software developers to ensure that their software is compliant with the emerging FAIR for research software principles¹⁹, and the community standards for FAIR nanoinformatics software being proposed herein. In many cases, nanoinformatics models are designed to predict specific regulatory-relevant endpoints such as genotoxicity, acute or reproductive toxicity, neurotoxicity etc., with the goal of replacing the experimental tests, and thus there is a requirement for validation for regulatory use, a critical step of which is documentation of the prediction principle or hypothesis, and documentation of the underpinning dataset(s). Section 2 of the current document provides an overview of the current approaches for documentation of predictive toxicology models including QSARs and physiologically-based pharmacokinetics models, and emerging standards for physics-based models, and discusses how further implementation of these harmonised approaches will facilitate the FAIRification of nanoinformatics models and software.

¹⁸ Many journals accept software papers: <https://the-turing-way.netlify.app/communication/dif-articles/software.html>

¹⁹ <https://rd-alliance.org/group/fair-research-software-fair4rs-wg/outcomes/fair-principles-research-software-fair4rs-0>

2. Community standards for documentation of *in silico* approaches for (nano)safety assessment

Addressing the challenges arising from the increasing use of machine learning and Artificial Intelligence needs to begin with clear and harmonised documentation (FAIRification) of the models and the corresponding training and validation datasets. Additionally, the process of validation and then the obtained results also need to be documented and made FAIR. Since QSAR models are well established for chemicals, and indeed many have been approved for use in chemicals' risk assessment, there is already a well-established process for validation and documentation of QSAR models via the QSAR model report form. In contrast, use of physics-based models in risk assessment of materials safety is more recent and thus the formalisation of data and model reporting processes is less established, but a community developed pre-standard has been developed and is increasingly being completed, called MODA, or Modelling and Data. These model documentation approaches (QMRF and MODA), which can be considered as metadata for the related models, are described in detail below. Their role in FAIRification of research software, and as a means for increasing the FAIRness of the models themselves and the software resulting from them, is then presented.

2.1. QMRF templates as a means of documenting data-driven *in silico* models

QSAR models are mathematical models used in chemistry and pharmacology to predict a wide range of properties such as biological activity, environmental effects, and physicochemical characteristics of various chemical compounds. (Q)SAR/(Q)SPR modelling is a non-animal testing method applied to predict the physicochemical properties, biological activity, (eco)toxicological and environmental fate properties of substances including nanomaterials. The (Q)SAR/(Q)SPR modelling paradigm is based on the relationships between chemical structures and observed (experimentally measured or theoretically simulated) properties of substances, where the chemical structure is represented in a numerical manner by so-called 'descriptors'. Examples of these descriptors include 'periodic table descriptors' such as ionisation energy, the energy needed to remove a metal ion from the surface of a nanomaterial. These descriptors are more akin to the physico-chemical characterisation data associated with real nanomaterials samples, and as such as are complementary to chemical representation notations such as InChI and SMILES, which describe the interconnectivity of the atoms and their spatial arrangements.

(Q)SAR/(Q)SPR models are built on the basis of limited datasets for chemicals, and the identified relationships can be used to fill wide data gaps for substances for which a modelled parameter is not available. The reliability of the (Q)SAR/(Q)SPR is proven with a series of statistical parameters and the credibility of predictions for new substances depends on their presence in the applicability domain of the model. To support regulatory adoption of QSAR models, the OECD member

countries agreed on the principles for validating (Q)SAR models²⁰ for their use in regulatory assessment of chemical safety. The core of the validation process is the QMRF (QSAR Model Reporting Format), a standardised reporting guideline for QSAR models. The QMRF (Figure 3) ensures that the models have undergone rigorous internal and external validation processes, making their predictions more reliable. This is particularly important when the models are used to make safety assessments for new drugs or environmental chemicals (and increasingly, for nanomaterials).

Detailed guidance on the steps in the design, validation and documentation of computational approaches such as QSARs¹⁷, physiologically based pharmacokinetics (PBPK) models²¹ and integrated approaches such as IATAs²² are available from the OECD and the European Chemicals Agency (ECHA) for chemicals and nanomaterials (NMs), and the well-established processes for the validation of alternative test methods via EURL ECVAM, as described in Worth and Balls (2004) also apply in principle to chemoinformatics and nanoinformatics models.

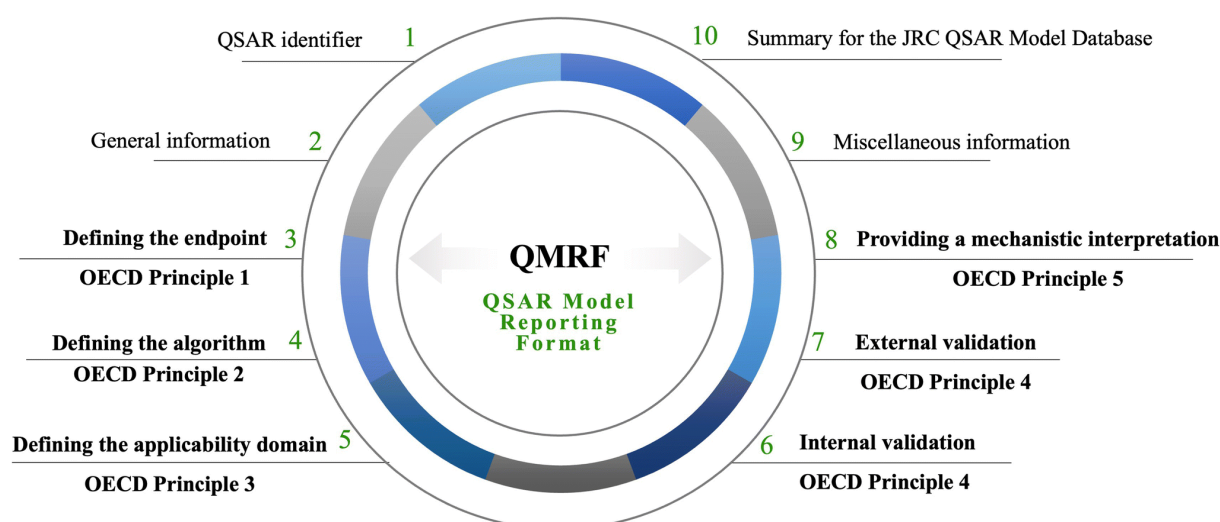


Figure 3: The structure of the QMRF document. Each of the sections are explained below. From Sosnowska et al., 2023.

²⁰ Organisation for Economic Co-operation and Development. Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. 1–154 (2007)

https://read.oecd-ilibrary.org/environment/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models_9789264085442-en#page1

²¹ OECD. Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes. Series on Testing and Assessment No. 331. <https://www.oecd.org/chemicalsafety/risk-assessment/guidance-document-on-the-characterisation-validation-and-reporting-of-physiologically-based-kinetic-models-for-regulatory-purposes.pdf>

²² OECD. Report On Considerations From Case Studies On Integrated Approaches For Testing And Assessment (IATA). (2023) [https://one.oecd.org/document/ENV/CBC/MONO\(2023\)31/en/pdf](https://one.oecd.org/document/ENV/CBC/MONO(2023)31/en/pdf)

The QMRF is divided into multiple sections as shown schematically in Figure 2, each designed to address specific aspects of model quality ranging from basic identifiers and general information to highly detailed sections on algorithm definition, statistical validation, and endpoint definition:

- **Sections 1 and 2:** These initial sections provide an overview of the model, laying the groundwork for more technical details. This includes what the model aims to do, who developed it, and what its foundational elements are.
- **Sections 3 to 5:** These sections are more technical, detailing the endpoint (what you're trying to predict) and the algorithm (how you're trying to predict it), as well as the domain of applicability (under what conditions the model is valid).
- **Sections 6 and 7:** These are heavily focused on statistical validity, ensuring that the model not only works theoretically but has been tested for its reliability in predictions.
- **Section 8:** This focuses on the mechanistic interpretation, offering scientific reasoning behind why the model works, based on chemical and biological principles.
- **Sections 9 and 10:** These wrap up any other essential details, including miscellaneous information and how the model fits into larger databases or repositories.

By fulfilling each section of the QMRF, researchers ensure that their QSAR/QSPR models meet the highest standards of scientific rigour, making them more likely to be accepted by regulatory agencies for risk assessments and other purposes. Indeed, lack of appropriate documentation has been cited as the leading cause for rejection of QSARs. All QSAR models developed within the NanoCommons, NanoSolveIT and CompSafeNano projects, coordinated by the co-leads of WorldFAIR WP04, are accompanied by a QMRF report, that is part of the materials included in the web-application accompanying the model, and in the Supplemental File of the publication describing the model and its domain of applicability. An example of this best practice is provided in Section 3.

2.2. MODA templates as a means of documenting physics-based materials models

MODA, which stands for "Modeling and Data" is a template increasingly used in materials modelling projects to provide a comprehensive description of the modelling and simulation processes, as well as for data management. The idea is to present not just what models are used but also how and why they are chosen, how they are connected, and how the simulations are run. The MODA framework encapsulates the full lifecycle of a modelling and simulation study, from problem definition to computational details to post-processing, allowing for standardised, transparent, and reproducible research.

Significant effort has been devoted in the last few years to the development of predictive models for nanomaterials properties, effects and impacts in relation to their short- and long-term

effects and to facilitate the read-across from data-rich nanomaterials to data-poor ones. However, much of this focus, and the focus of the regulatory validation processes presented above, has been on so-called data driven models and predictions of toxicity effects, rather than on models predicting materials properties or interactions, often called materials modelling.

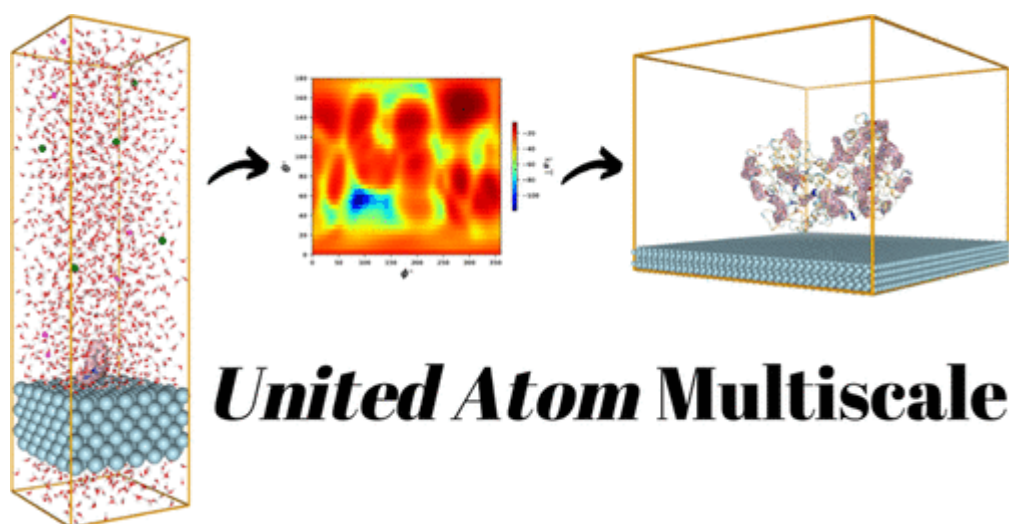


Figure 4: UnitedAtom (UA) is a bottom-up multiscale approach to estimate protein adsorption energies onto NMs, whereby globular proteins are represented as a collection of one-per-amino-acid coarse-grain (CG) beads. The total interaction potential between the protein and NM is evaluated by summation of all types of molecular interactions between the CG beads and the NMs. The total UA adsorption energy of the protein onto NM has two major contributions arising from the short-range forces, computed with high resolution all-atom MD simulations, and the long-range forces, evaluated via CG potentials. From Subbotina et al., 2022.

To address this gap, the report on materials modelling published by the European Commission (European Commission, 2017) provided a detailed description of “physics-based models”, the scales at which they operate, and the corresponding classes of models, i.e., electronic, atomic, mesoscopic, continuous. Within NanoCommons, NanoSolveIT and CompSafeNano, much physics-based modelling work has spanned these scales, as shown schematically in Figure 4 for the various scales required to model the interactions of proteins with nanomaterials (Subbotina and Lobaskin, 2022). The protein corona, or the layer of biomolecules that binds to nanomaterials and converts their original synthetic identity into an acquired biological or environmental identity, depending on the nature of the surroundings, is a critical determinant of nanomaterials uptake, processing and biodistribution in organisms (Lynch and Dawson, 2008; Wheeler et al., 2022). In order to computationally predict the acquired protein corona, which determines the subsequent fate and effects of the nanomaterials in living systems, requires the combination of coarse-grained approaches and molecular dynamics (MD) simulations to represent the long-range and short-range interactions respectively (Subbotina et al., 2022). Describing these models via the QMRF approach is not possible and thus it was noted by the European Materials Modelling Council and others that a reporting framework for materials modelling / physics-based models was missing. To overcome

this gap, the materials modelling terminology, classification and metadata were developed, and the resulting Modeling and Data (MODA) template was standardised via a CEN Workshop Agreement in 2018 ([CWA 17284](#)²³).

The purpose of MODA is to make it as clear as possible what was done, why, and how, so that other researchers can understand, validate, or build upon the work. This is particularly important in materials science and engineering where complex simulations often involve multiple models, each with its assumptions, and where the data must often be post-processed in some way to be useful.

The structure of a MODA document is designed to standardise the description, documentation, and communication of computational models and simulations. Figure 5 summarises the elements of physics-based models that are captured in a MODA report, which can be considered as serving an equivalent purpose as the QMRF does for documentation of data-driven models. While the exact structure can vary depending on the specific scientific field, the application, or the complexity of the model, a typical MODA framework may consist of the key sections presented in Table 1 below.

The importance of MODA lies in its ability to make complex modelling projects more transparent, understandable, and reproducible, thus advancing the quality and impact of research in materials science and related fields.

An interesting feature of MODAs is that they can be used to describe linked models, and indeed the MODA CWA provides a schematic (see Figure 6) of workflow templates for a stand-alone model, for a chain of linked models, for a chain of iteratively coupled models and for tightly-coupled models.

²³ Comité Européen De Normalisation (CEN) Workshop Agreement - CWA 17284 - Materials modelling - Terminology, classification and metadata. (2018) https://www.cenelec.eu/media/CEN-CENELEC/CWAs/RI/cwa17284_2018.pdf

HEADING ^a	
OVERVIEW OF THE SIMULATION ^b	
WORKFLOW ^c	
Simulation with Model 1 ^d	1. Aspect of the user case/system to be simulated
	2. Governing equations
	3. Solver and computational translation of the specifications
	4. Post-processing
Simulation with Model 2	1. Aspect of the user case/system to be simulated
	2. Governing equations
	3. Solver and computational translation of the specifications
	4. Post-processing
...	...
Simulation with Model N	1. Aspect of the user case/system to be simulated
	2. Governing equations
	3. Solver and computational translation of the specifications
	4. Post-processing
^a Heading, including name of the user case, project, owner ^b Overview of the simulation, including the chain of models used ^c Workflow, i.e. a graphical representation of the simulation ^d Description of each part of the simulation pertaining to one model used in the chain	

Figure 5: The core elements of the MODA template. Notably each model is described with the four key sections - the aspect being modelled, the equation, the solver and computational specifications and any post-processing (see Table 2 for further details of the constituents of each section), and for multi-models the relationships between the models, and the data-flow between them is described as schematically as shown in Figure 6. From CEN CWA 17284.

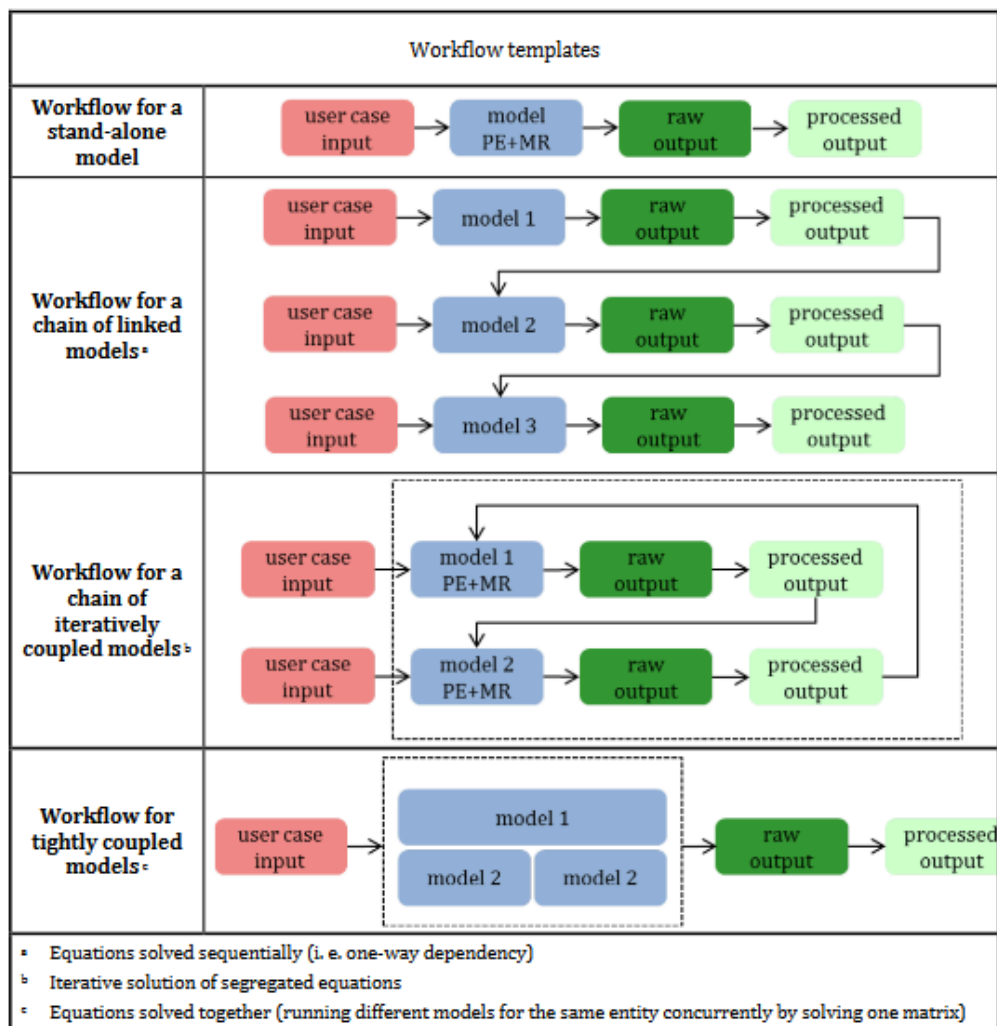


Figure 6: MODA enables clear visualisation of the relationships between models, and the data-flow between them through workflows for individual models, chains of linked models, chains of iteratively coupled models or tightly-coupled models. From CEN workshop agreement CWA 17284.

2.3. Role of MODA in standardising modelling and simulation for nanomaterials

In the field of nanomaterials research, modelling and simulation are invaluable tools for understanding material properties, behaviours, and mechanisms at the atomic or molecular scale. However, the highly interdisciplinary nature of this research, involving physics, chemistry, engineering, and computational science, can lead to a lack of standardisation and comparability between different studies. MODA plays a significant role in enhancing the rigour, transparency, and collaborative potential of modelling and simulation studies in nanomaterials research. By providing a standardised framework, MODA helps to elevate the quality and impact of research outputs in this rapidly evolving field.

Table 1: The elements of the MODA template for reporting on physics-based models to support building of regulatory acceptance. Adapted from NanoSolveIT deliverable report D6.4 (Report on outcome of international Round Robin for model benchmarking) which included development of a Standard Operating Procedure for validation of *in silico* models and predictions on the properties, exposure, fate, and toxicity or hazard of nanoscale materials (NMs), as a means to guide model developers through the validation processes and to increase the regulatory acceptance of *in silico* new approach methodologies as part of an Integrated Approach to Testing and Assessment (IATA).

Purpose of Section	Elements of the Section	Why These Elements are Important
<p>1. User Case</p> <p>What is the real-world scenario or problem that the simulation aims to address?</p> <p>Is the mode stand-alone or a series of models used to transform initial input data into the final output, to address the User Case?</p> <p>Information about any publications that validate or peer-review the data being used or generated.</p>	<p>Aspect of the User Case to be Calculated: Specifies what the model aims to address or simulate.</p> <p>Material: Defines the material or system being modelled.</p> <p>Geometry: The physical dimensions or layout of the system.</p> <p>Time Lapse: Specifies simulation time frame.</p> <p>Manufacturing Process or In-Service Conditions: Conditions under which the material or system is manufactured or operated.</p> <p>Publication on this One Data-mining Operation: Any relevant publications or references.</p>	<p>Unified Terminology: MODA provides a common language for researchers to describe their models, including the equations used, the assumptions made, and the limitations inherent in the model.</p> <p>Consistent Documentation: Using MODA templates ensures that all key elements of a model (physics equations, materials relations, and computational aspects), are consistently documented. This helps others to understand, replicate, or extend the work.</p> <p>Transparency: MODA templates provide a clear and organised way to display all the important aspects of a modelling project, ensuring that the models, methods, and data can be fully understood by others.</p>
<p>2. Generic Physics of the Model Equation</p> <p>An explanation for why this combination of models and particular workflow were chosen.</p> <p>Describes how exactly the models are implemented computationally, including</p>	<p>Model Type and Name: Classification and name of the physical equation (PE).</p> <p>Model Entity: Describes the scale at which the model operates (atoms, grains, finite volumes).</p>	<p>Common Ground: Nanomaterials research often involves experts from different fields. MODA provides a common platform for these experts to collaborate effectively, as it ensures that everyone is "speaking the same language." Currently, much of the MODA is free-text based, and the materials modelling ontology provides some guidance but as an</p>

<p>things like the numerical methods used, software tools, and solver parameters.</p>	<p>Model Physics/Chemistry Equation: Lists and describes the equations that the model solves.</p> <p>Materials Relations: Any material-specific parameters or relations.</p> <p>Simulated Input: Describes the input used for the simulation, and how it is calculated.</p>	<p>upper ontology addresses the concepts at a high level of granularity. Work within WorldFAIR is underway to develop a web-interface and structured template for MODA to reduce the need for free text inputs and to support harmonisation.</p>
<p>3. Solver and Computational Translation of the Specifications</p>	<p>Numerical Solver: Type and name of the numerical method used for solving the model equations.</p> <p>Software Tool: The software used for simulation.</p> <p>Time Step: Numerical time step for the simulation.</p> <p>Computational Representation: How the physical equations and material relations are computationally implemented.</p> <p>Computational Boundary Conditions: Boundary conditions in the computational domain.</p> <p>Additional Solver Parameters: Any additional parameters like tolerances, cut-offs, convergence criteria, etc.</p>	<p>Building upon Existing Work: The transparent and standardised documentation allows other researchers to build upon existing models or to integrate them into larger, more complex systems.</p> <p>Feedback Loop: MODA's comprehensive structure allows for a better feedback loop between experimentalists and theoreticians, allowing each to inform the work of the other more effectively.</p>
<p>4. Post-Processing</p> <p>What steps are taken after the raw data is generated by the simulations, and why?</p>	<p>The Processed Output: Describes what kind of output is obtained after post-processing.</p> <p>Methodologies: Explains the post-processing methods used.</p>	<p>Critical Evaluation: By outlining the methodology in detail, MODA enables reviewers and other researchers to critically evaluate the quality and reliability of the modelling study.</p> <p>Benchmarking: The framework facilitates comparison of different</p>

<p>An assessment of the reliability of the models and the simulated data.</p>	<p>Margin of Error: Defines the reliability and error bounds of the simulation results.</p> <p>For data-based models, include sections that describe:</p> <p>Equation Type and Name: Classification of the data-based model (e.g., machine learning model, statistical model).</p> <p>Database and Type: Source and type of the data used.</p> <p>Equation Hypothesis: The assumptions or hypotheses behind the data-based model.</p> <p>Physical Quantities: Parameters and variables that appear in the data-based model.</p>	<p>models and methods by providing a standardised means to describe them, which is particularly useful for developing benchmarks in nanomaterials research.</p> <p>Data Management: Include details on data sources, storage, and accessibility assists in data management, which is crucial for large projects and is increasingly a requirement of research funding agencies.</p>
<p>5. Computational Detail of Data-mining Operation</p> <p>Details on whether the model and/or data are open source, commercial, or otherwise, and how they can be accessed.</p>	<p>Numerical Operations: Describes the numerical methods used in the data-mining.</p> <p>Software Tool: Software used for data-mining.</p> <p>Margin of Error: Reliability and error bounds of the data-based model.</p>	<p>Transparent Workflow: MODA outlines the modelling workflow from input data to computational methods and post-processing, thus facilitating reproducibility.</p> <p>Data Sharing: MODA encourages researchers to detail the data sources and conditions under which data can be accessed, which is crucial for validating and comparing results.</p>

2.4. Role of QMRF/MODA in increasing the FAIRness of nanoinformatics models

The previous sections have demonstrated the importance of the QMRF, and the growing importance of the MODA, as vital frameworks for standardising the presentation of data-driven (QSAR models, machine learning models) and physics-based materials models, respectively, and their use for documentation of nanoinformatics models. These reporting templates ensure that the models are transparent, reliable, and are essential steps towards formal validation, which is of

paramount importance for their acceptance in regulatory decisions concerning nanomaterials. By presenting data in a standardised, transparent, and validated format, QMRF and MODA facilitate scientific discovery and help to build researcher and public trust. The ethical implications of introducing new materials into various sectors, especially healthcare and the environment, make this trust exceptionally crucial. Furthermore, the QMRF and MODA reporting frameworks foster collaborative efforts to improve existing models, extend their domains of applicability, integrate and link models - via consensus models, meta-models or through integration into an Integrated Approach to Testing and Assessment (IATA). By adhering to a standardised reporting structure, a common approach to data and knowledge sharing, model integration becomes more straightforward, and the chances of getting the models validated and approved for regulatory use increases.

Beyond these advantages in supporting regulatory transparency, QMRF and MODA templates can also help to increase the FAIRness of the models themselves. Thus, the QMRF and MODA templates can be considered to be FAIR Enabling Resources (FERs) in the context of a FAIR implementation profile (FIP) (see also WorldFAIR Deliverable report D4.1 for further details of the FIP developed for nanomaterials safety), although we note that the current version does not yet include QMRF and MODA, but the final version due in April 2024 will.

Table 2 maps the FAIR4RS principles to QMRF and MODA and demonstrates that these templates can be considered as community standards, and as metadata schema and provide provenance information and unique and persistent identifiers for the models they describe and accompany.

Table 2: Analysis of the role of QMRF and MODA templates for describing nanoinformatics models in enhancing the FAIRness of models.

No.	FAIR4RS Principle description	QMRF approach	MODA approach
F1.	Software is assigned a globally unique and persistent identifier.	Yes - each model has a specific name and a link to its web-application.	Yes - each model as a specific name and a link to its web-application
F1.1.	Components of the software representing levels of granularity are assigned distinct identifiers.	Yes - each model has a unique number/identifier and the sequence they are linked in is also captured based on the output and input information for each model (see Figure 5).	Yes - each model has a unique number / identifier and the sequence they are linked in is also captured based on the output and input information for each model (see Figure 5).
F1.2.	Different versions of the software	Versioning is tracked,	MODA captures version of

	are assigned distinct identifiers.	though may not be as explicit as MODA.	software used. MODA templates can themselves be versioned.
F2.	Software is described with rich metadata.	Detailed metadata are provided for each model, capturing its specifics.	MODA is the means to provide the metadata
F3.	Metadata clearly and explicitly include the identifier of the software they describe.	The identifier is included within the QMRF documentation.	Explicitly included in MODA documentation.
F4.	Metadata are FAIR, searchable and indexable.	Metadata structured to align with FAIR principles.	Designed for FAIR compliance; easily searchable and indexable.
A1.	Software is retrievable by its identifier using a standardised communications protocol.	Typically retrievable via web-based applications.	MODA supports retrieval of model information through standard protocols.
A1.1.	The protocol is open, free, and universally implementable.	Depends on the specific model and platform used.	Generally adheres to open and universal protocols where possible.
A1.2.	The protocol allows for an authentication and authorization procedure, where necessary.	Authentication and authorization procedures are model and platform-dependent.	MODA documentation supports inclusion of such procedures where applicable.
A2.	Metadata are accessible, even when the software is no longer available.	Metadata availability is maintained independently of software availability.	Ensured through MODA's comprehensive documentation approach.
I1.	Software reads, writes and exchanges data in a way that meets domain-relevant community standards.	Varies with the model; generally designed to meet community standards.	MODA encourages adherence to domain-specific data standards.
I2.	Software includes qualified references to other objects.	Includes references as part of the model documentation.	Systematic inclusion of references to other objects in the documentation.
R1.	Software is described with a plurality of accurate and relevant	Attributes are detailed	Comprehensive description

	attributes.	within the QMRF framework.	is a key component of MODA.
R1.1.	Software is given a clear and accessible licence.	Licence details are included where applicable.	Licensing information is explicitly stated in MODA documentation.
R1.2.	Software is associated with detailed provenance.	Provenance is captured as part of the model's development history.	Detailed provenance is a core aspect of MODA documentation.
R2.	Software includes qualified references to other software.	References to other software are included where relevant.	Systematic referencing of related software.
R3.	Software meets domain-relevant community standards.	Designed to align with community standards in nanoinformatics.	MODA framework ensures adherence to community standards.

Given the status of QMRFs in regulatory assessment, there is already an online editor to support their creation in a harmonised manner, via the QMRF editor, which can be downloaded from [Sourceforge](#)²⁴. No such editor exists as yet for MODA, and as a final activity within WorldFAIR we will develop a web application to support streamlined MODA completion, and as part of a registry of nanoinformatics models. Full details on this will be provided in the final WorldFAIR WP04 deliverable report D4.3, 'Nanomaterials Human and machine-readable provenance and persistence policies'.

3. Recommendations and tools for FAIR nanoinformatics models

Based on our review of the field of FAIR for software, the experience from the broader chemoinformatics and *de novo* drug design fields, and the extensive work performed in the NanoCommons and NanoSolveIT projects to democratise access to *in silico* nanoinformatics approaches and support their transition into regulatory use, we have distilled out the following core actions needed to enhance the FAIRness of nanoinformatics models. Most of these actions are the responsibility of the model and software developers, i.e., documentation of the models, provision of rich metadata related to the model / software, provision of access protocols such as a web interface, an application programming interface (API) or a KNIME workflow, registration of the model / software to the community registry, while others are the responsibility of the broader

²⁴ QMRF templates can be completed via a html editors available at Sourceforge: <https://qmrf.sourceforge.net/>

nanoinformatics community, as represented by WorldFAIR WP04, such as the implementation of a registry for nanoinformatics models, workflows for data curation and quality assessment to support and underpin model and software development, and for re-integration of the predicted data into the database to ensure its ongoing development. The following subsections present these recommendations in further detail.

3.1. Fully documented models / FAIR models

In conjunction with the OECD Test Guidelines²⁵ that describe how experimental toxicity data is generated, the integration of QMRF and MODA templates offers a systematic way of documenting *in silico* models. The QMRF template is particularly useful for detailing data-driven models, such as QSAR models. It enables researchers to concisely report on the methodology, development, validation, and intended application of their models. This level of detailed documentation is essential for transparency, allowing other researchers and stakeholders to fully understand the model's basis and limitations. Similarly, the MODA template is instrumental for documenting physics-based materials models. It provides a standardised format to describe various aspects of the modelling process, including the theoretical basis, computational details, and the specific algorithms used. The use of MODA templates ensures that all critical information about the model's development and application is comprehensively recorded, facilitating replication and further research. The combination of OECD Test Guidelines with QMRF and MODA templates forms a robust framework for the development, validation, and documentation of *in silico* models in nanoinformatics. By following these established protocols and templates, researchers can ensure that their models are not only scientifically sound but also well-documented and transparent. This is particularly important in a field that relies heavily on computational methods to guide decision-making and regulatory approvals. Such a meticulous approach to validating and documenting models enhances the reliability and acceptance of *in silico* methods in the scientific community and regulatory bodies, thereby advancing the field of nanoinformatics.

The [NanoSolveIT](#) project developed a Standard Operating Procedure (SOP)²⁶ to support researchers in documentation of nanoinformatics models, based on the well-established steps in model development and supplemented with the best practice developed within NanoSolveIT on documentation of models, on integration of models as meta-models, consensus models and/or as an IATA, on deployment of models as web-services, with programmatic access via Application programming interfaces (APIs) and through computational platforms such as [Jagpot](#)²⁷, [Isalos Analytics Platform](#)²⁸ and others. We extend this with information and best practice extracted from the analysis of the processes for formal regulatory validation of alternative test methods for regulatory use (by ECVAM) which applies also to *in silico* methods, and the recent OECD

²⁵ [OECD Test Guidelines for Chemicals - OECD](#)

²⁶ Article in preparation for special issue of Beilstein Journal of Nanotechnology, Submission March 2024.

²⁷ <https://www.jagpot.org/>

²⁸ <https://isalos.novamechanics.com/>

publication on a framework for QSAR model validation²⁹ which provides details on the elements to be assessed and maps these to the specific points of the QMRF and the QSAR model prediction form (QPRF). As far as possible, the steps for data-driven models, physics-based models and combined models (multi-models, consensus models, IATA) are harmonised, and mapped to the respective reporting frameworks (QMRF and MODA). Thus, the steps of the NanoSolveIT nanoinformatics modelling SOP, extended via the FAIR4RS checklist as shown in Table 2, as illustrated schematically in Figure 7, include:

1. Compilation of the data and inputs on which the model will be developed (e.g., curated datasets for data-driven models, or for example crystallographic information files (in CIF format) for physics-based models, and alignment of the outputs and inputs for linked or meta-models);
2. Development and testing of the model (e.g., data splitting, and checking of prediction for machine learning (ML) models; comparison with experimental results for physics-based models, a combination approach for meta-models and integrated models);
3. Validation of the model, according to the OECD principles for QSAR models (e.g., using relevant statistics to test the predictive power and define the domain of applicability, while for physics-based models, the model is validated through experimental verification of the prediction or comparison to literature values, with a multi-step verification for meta-models).
4. Documentation of the model, its underpinning approach and its predictive power, via the QMRF for data-driven models and the MODA for physics-based ones, and a combination for meta-models. Additional documentation to support application of the models should include development of training materials, example datasets etc.). This step addresses FAIR4RS principles F1, F2 and A2, as shown in Figure 7.
5. Deposition of the underpinning datasets, and the predictions, into a recognised repository to enable reuse and/or to feed into a subsequent model. Within NanoSolveIT, datasets were uploaded to the NanoSolveIT Knowledge Base and/or into the NanoPharos database which provides ready-for-modelling datasets, both of which allow interaction with the data, and archived via Zenodo, addressing FAIR4RS principles F3, F4 and A2, as shown in Figure 7.
6. Deployment the models as web applications or enable user access via computational notebooks or other approaches. A key element of model uptake is accessibility and access. This step addresses FAIR4RS principles A1, A1.1, A1.2, I1, R1, R1.1, R1.2 and R3, as shown in Figure 7.

²⁹ (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure – Activity Relationship models, predictions, and results based on multiple predictions. Series on Testing and Assessment No. 386. [https://one.oecd.org/document/ENV/CBC/MONO\(2023\)32/en/pdf](https://one.oecd.org/document/ENV/CBC/MONO(2023)32/en/pdf)

7. Submission to ECVAM of a completed pre-submission enquiry template (via email) via a Test pre-submission form (TPF) - see Figure 4 and Annex 2 for complete document - which allows ECVAM to undertake a preliminary assessment of the suitability for formal validation of a Test Method, in this case an *in silico* nanoinformatics model or IATA.
8. If invited by ECVAM, submit the Test Submission Template to start the formal validation process. The template is provided by ECVAM upon invitation.
9. Follow the validation progress via the EURL ECVAM Tracking system for Alternative Methods towards regulatory acceptance ([TSAR](https://tsar.jrc.ec.europa.eu/)³⁰).
10. If needed, update the documentation and web-application based on feedback from ECVAM.

We note that here that there is a slight distinction between the underpinning model itself (the steps in the development of which is shown in in blue outlined boxes in Figure 7) and the resulting software that is being made FAIR - the software is the deployment of the model via a web application, as a KNIME workflow or provides programmatic access via an API, indicated by orange outlined boxes in Figure 7. The FAIRification steps are indicated via green box outlines - and constitute the critical steps of model documentation, data and metadata deposition and deployment of the model as a software, as shown in Figure 7.

³⁰ <https://tsar.jrc.ec.europa.eu/>, the system for tracking the progress of alternative, non-animal chemical testing methods through the validation process - it documents the stages methods have reached in terms of acceptance as a recognised test method for use in various sectors together with a summary description. Where available, TSAR includes relevant records and documents associated with a method linked to the different steps of the entire process: [submission](#), [validation](#), [peer-review](#), recommendations and regulatory acceptance including international standards represented in the tracking system.

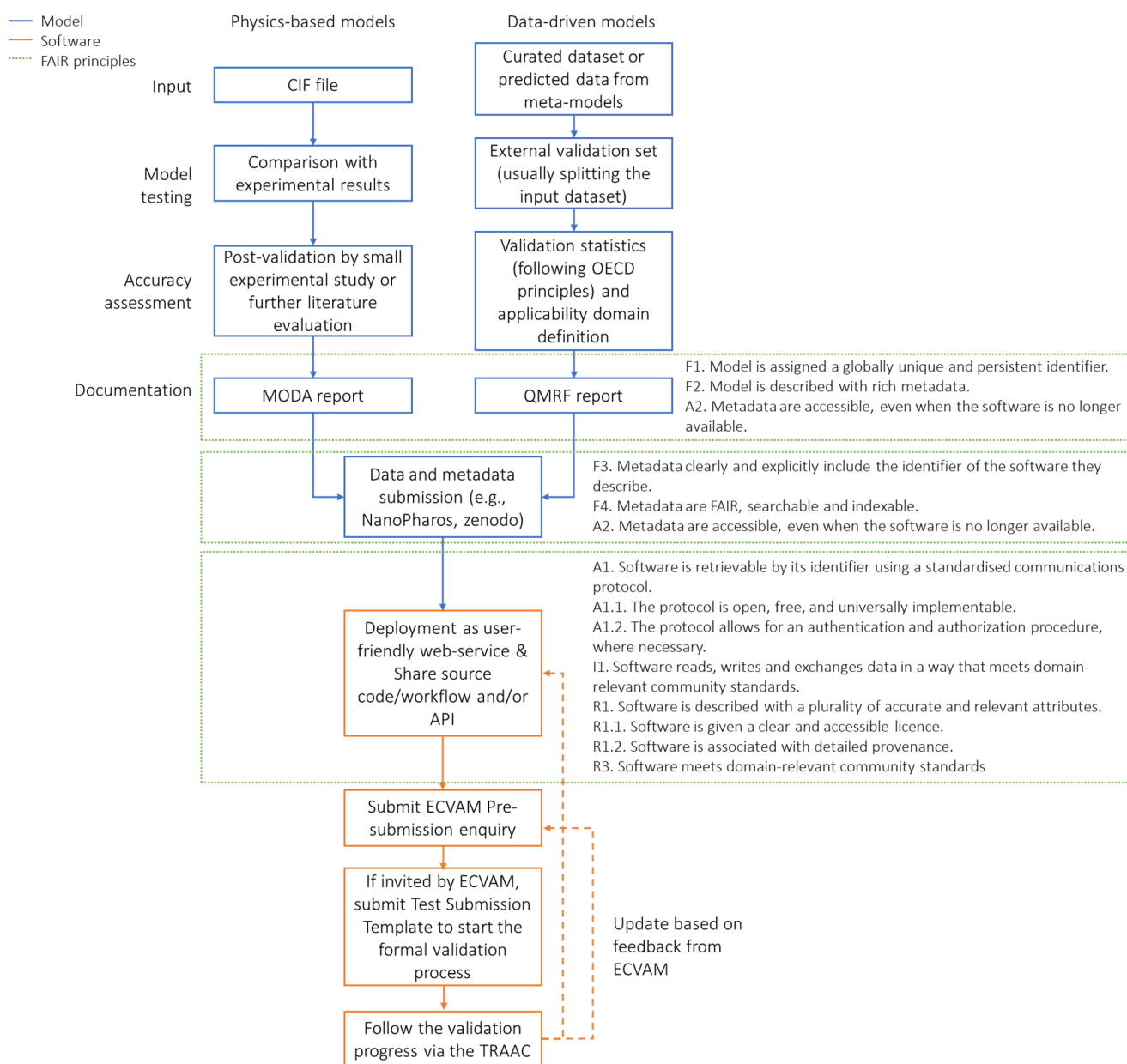


Figure 7. Schematic of the steps covered in the NanoSolveIT SOP for nanoinformatics model development and validation workflow extended to indicate how these also support FAIRification of the models and the resulting software (e.g., as web interfaces, KNIME workflows and/or APIs). Adapted from NanoSolveIT D6.4. Submitting the model / software to ECVAM is not part of the FAIRification process but is included for context, as it is the ultimate aim for much of the nanoinformatics software.

3.2. Rich metadata for nanoinformatics models

The role of metadata is multifaceted, and is more than just data about data; it serves as a critical guide that provides in-depth insights into the nature and characteristics of various nanoinformatics models. This metadata encompasses a range of information that is essential for users to understand and effectively utilise these models in design and *in silico* testing of nanomaterials and resulting products optimised for safety, sustainability and functionality. At its core, metadata includes specifics about the model's purpose – the 'why' behind its creation. This could range from understanding the behaviour of nanomaterials under certain conditions to predicting their interactions at a molecular level. By clearly defining the model's objectives, metadata helps users ascertain whether a particular model aligns with their research or regulatory needs.

The design and development aspect of the metadata delves into the 'how' of the model. This includes details about the theoretical frameworks, algorithms, and computational methods employed in building the model. Such information is crucial for users to assess the scientific robustness and validity of the model. It also provides insights into the underpinning scientific principles, allowing users to gauge the applicability of the model to their specific research scenarios.

Version history in the metadata is particularly important in a field that is rapidly evolving. It tracks the changes and updates made to the model over time, offering users a chronological map of its development. This aspect of metadata is crucial for understanding the evolution of the model's capabilities and for ensuring that the latest and most accurate version is being utilised. GitHub is one of the most useful tools to support model versioning for many model types, but we note that it was not developed with ML in mind, and the differences between software development and machine learning solution development and deployment, as shown in Figure 8, limit the usefulness of Git for ML models. Indeed, ML model development often requires tracking not just code, but also large datasets, model weights, hyperparameters, and experimentation results.

Usage instructions form another critical component of metadata. They provide guidance on how to operate the model, interpret its outputs, and integrate it into broader research workflows. This aspect of metadata is key to enabling users, regardless of their technical expertise, to effectively engage with the model. Clear and concise usage instructions also enhance the accessibility of the model, making it more approachable for a diverse user base. Similarly, the usage conditions and licensing must be clearly indicated as part of the model metadata, ideally using the Creative Commons licensing framework³¹.

In summary, transparency is achieved through the open disclosure of the model's development process and capabilities. Reproducibility is facilitated by providing detailed information that allows

³¹ <https://creativecommons.org/share-your-work/cclicenses/>

other researchers to replicate the model's results. Finally, reliability is bolstered by offering a thorough understanding of the model, its applications, and its limitations.

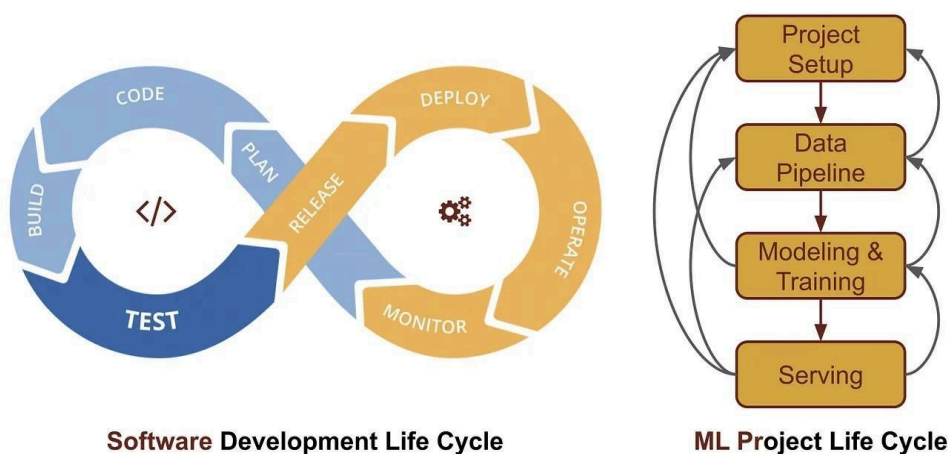


Figure 8: Differentiation between software development life cycle and ML project life cycle. From Gupta (2022)³².

3.3. Registry of nanoinformatics models

A registry of nanoinformatics models represents a critical infrastructure within the field, functioning as a centralised repository that meticulously catalogues an extensive range of models related to the various aspects of nanomaterials modelling. As such a registry was identified as a major gap for this community, WorldFAIR WP04 is implementing a nanomaterials modelling registry and will crowd-source models for inclusion. This registry is intended to be more than just a simple database; it will be a comprehensive resource that systematically organises and provides detailed information on various nanoinformatics models, their modelling principle, the end-point modelled, and the domain of applicability - in this case, typically the range of nanomaterials that it has been tested on. The significance of a registry of nanoinformatics models thus lies in its ability to streamline the process of model selection and application for researchers and practitioners.

At its core, the registry is designed to be highly searchable, allowing users to efficiently locate models that are relevant to their specific research needs or interests (see Figure 9). This searchability is a key feature, considering the vast and diverse array of models available in nanoinformatics. Users can filter and search for models based on various criteria such as the type of nanomaterials, the properties being modelled, the underlying computational methods, or the application domains. This level of granularity in search functionality greatly aids researchers in pinpointing the exact models that suit their project requirements. Each entry in the registry typically contains a wealth of information about the respective model, based on the

³² <https://www.linkedin.com/pulse/machine-learning-vs-traditional-software-development-ml4devs-gupta/>

community-agreed metadata regarding the models (see sections 2.2-2.4 above on the QMRF and MODA templates for model documentation) and section 3.2 for metadata for nanoinformatics models). This includes a detailed description of the model's functionalities, outlining what the model can do and its potential applications. Furthermore, the registry provides specific details about the application areas of each model, offering insights into the contexts or scenarios where the model has been designed to perform optimally.

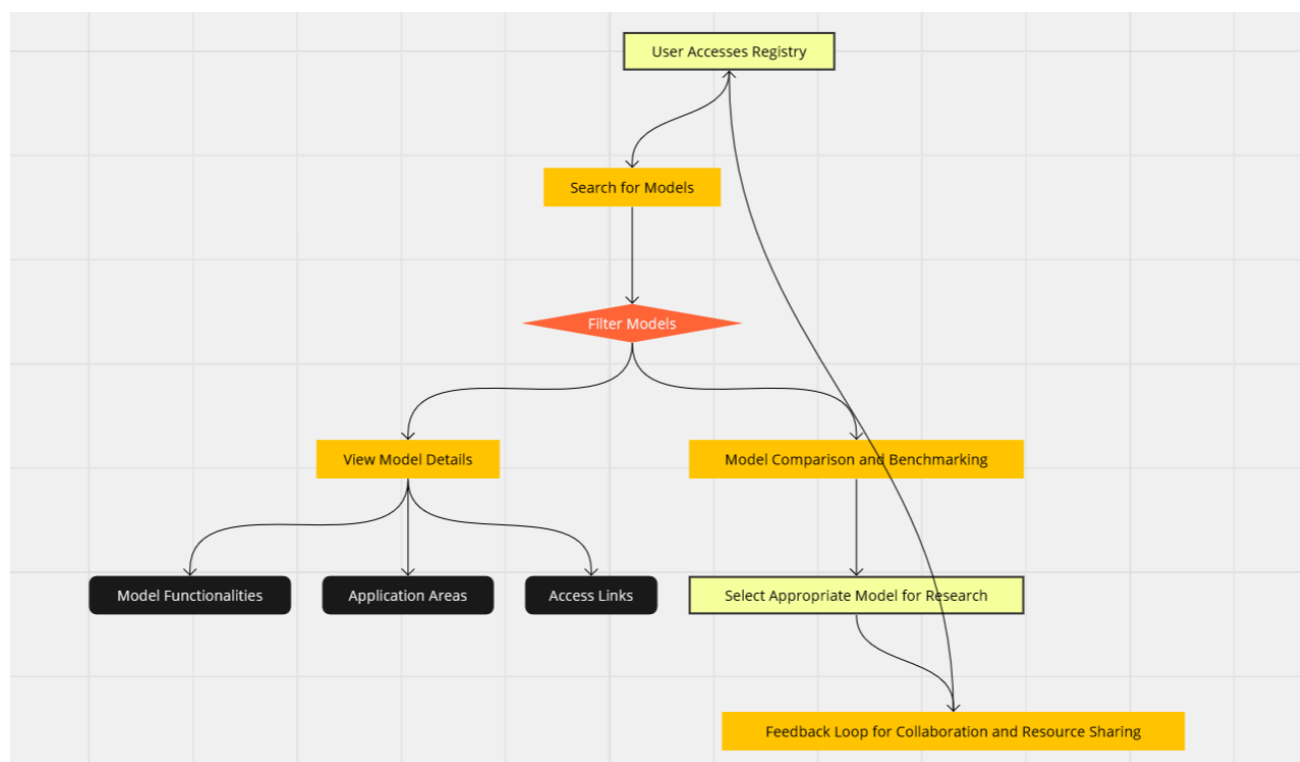


Figure 9: Schematic of the functions of the Nanoinformatics model Registry developed within WorldFAIR to address a gap identified in this research community in terms of maximising the FAIRness of nanoinformatics models and software. See also: https://miro.com/app/board/uXjVN98HjkQ=?utm_source=showme&utm_campaign=cpa.

Another critical aspect of the nanoinformatics registry is the inclusion of access links or instructions for each model. These links serve as direct gateways to the models, whether they are hosted on web platforms, available through APIs, or require specific software installations. The ease of access to models is instrumental in facilitating their use and integration into research workflows. The registry also fosters a collaborative environment within the nanoinformatics community. Provision of a shared platform where models are openly catalogued, is expected to encourage the exchange of resources and knowledge among researchers and institutions. This collaboration is pivotal in a field that thrives on interdisciplinary approaches and collective advancements.

Moreover, the registry can serve as a benchmarking tool, where the performance and utility of different models can be compared and analysed (see Figure 9). This benchmarking is essential for the continuous improvement of models and for guiding future developments in nanoinformatics modelling. The registry of nanoinformatics models is a dynamic and interactive tool, developed by WorldFAIR partner NovaMechanics Ltd.³³ and deployed via its not-for-profit sister organisation Entelos Institute³⁴, that is being developed via WorldFAIR, based on the needs analysis performed as part of the development of the WorldFAIR WP04 FIP³⁵. The nanomaterials models registry enhances the efficiency, accessibility, and collaboration within the field and plays a crucial role in supporting researchers by providing a structured and comprehensive overview of available models, thereby enabling informed decision-making and fostering advancements in nanoinformatics research.

3.4. Curation of datasets for data driven models

Data-driven models are fundamentally dependent on the underlying datasets' quality, organisation, and accessibility. The efficacy of these models is thus intrinsically linked to how well the datasets they utilise are curated, and how well the experiments were performed and documented, i.e., how complete their metadata is. Various approaches to assess nanomaterials safety dataset completeness and quality have been presented (e.g., Marchese Robinson et al., 2016; Fernández-Cruz et al., 2018) as well as more recently, tools to automatically assess the quality and completeness of dataset for regulatory purposes (e.g., Basei et al., 2022; Shao et al., 2023).

The quality of a dataset in this context refers to its comprehensiveness, accuracy, and relevance to the specific nanomaterials being studied. High-quality datasets are those that are free from errors, represent a broad spectrum of nanomaterials properties, and contain data points that are relevant and up-to-date. The organisation of these datasets entails structuring the data in a manner that aligns with the requirements of the models being used. This involves formatting the data to ensure compatibility with various computational tools and algorithms, and categorising it in ways that facilitate efficient data retrieval and processing.

Accessibility implies that datasets should not only be available to researchers and model developers but also be presented in a format that is easily interpretable and usable. This involves ensuring that datasets are stored in formats that are universally recognised and can be readily integrated into different nanoinformatics software and platforms. The standardisation of datasets is a multi-faceted process. It not only involves aligning data formats to industry or research community standards but also includes the implementation of consistent data labelling and metadata provision. As described in Section 3.2, metadata plays a crucial role, providing context to

³³ NovaMechanics Ltd.: <https://novamechanics.com/>

³⁴ Entelos Institute: <https://entelos.eu/>

³⁵ WorldFAIR FIPs are being collected at <https://zenodo.org/communities/worldfair-project-fips>

the data, such as information about its source, collection methods, and any preprocessing steps the data has undergone.

The [Enalos Cloud platform](#)³⁶ offers a range of tools for the calculation of molecular and atomistic descriptors, a process that is especially critical for the computational enrichment of experimental datasets for development of nanoinformatics models. The first step in this process, *data cleaning*, involves the removal of errors or inconsistencies from within the datasets. This could include the elimination of duplicate records, correction of typographical errors, or addressing missing values. It's a step that ensures the integrity and reliability of the data, which is crucial for any subsequent analysis.

Structuring the data is the next critical phase. This involves organising the data in a manner that is not only coherent but also aligns with the specific requirements of the modelling tools being employed (converting file formats or restructuring data tables).

Standardisation, another fundamental aspect of organising datasets for modelling, ensures that the data adheres to established norms and protocols, facilitating interoperability across different modelling platforms.

Databases like [NanoPharos](#)³⁷ play an integral role in this overall process. They provide a structured framework for storing and managing nanomaterials data, making them ideal repositories for datasets that need to be FAIR-compliant and providing the ability to efficiently retrieve, share, and utilise data, which can significantly accelerate research and discovery. Enalos Cloud and NanoPharos can thus enhance the utility and applicability of data-driven models in nanoinformatics. This dataset organisation process (including the development of the data and metadata schema) lays the foundation for the development of robust and reliable models, which are capable of providing meaningful insights into the characteristics and behaviours of nanomaterials. The process facilitates current analytical needs but also paves the way for future advancements in the field, fostering a more collaborative and productive research environment.

3.5. Automated integration of datasets into modelling tools via KNIME Nodes

The integration of datasets into modelling tools is a critical process in nanoinformatics, one that has been greatly facilitated by the use of KNIME¹⁰, an open-source platform for data analytics, reporting, and integration. KNIME's architecture, centred around the concept of nodes – which are essentially the basic processing units of the platform – enables a highly flexible and automated approach to data workflows. These nodes act as individual data processors that can perform a wide range of functions, from simple data manipulation tasks to complex analyses and integrations.

³⁶ Enalos Cloud Platform: <http://www.enaloscloud.novamechanics.com/>

³⁷ NanoPharos database for curated datasets to support nanoinformatics modelling: <https://db.nanopharos.eu/>

KNIME nodes enable researchers to directly integrate datasets from specialised repositories, such as NanoPharos, into various modelling software. This capability is crucial for several reasons. Firstly, it streamlines the process of data curation to model development and application. Once the data is curated into one such data repository and ready for analysis, KNIME nodes can be employed significantly reducing the time and effort typically required to manually prepare and input data into modelling software. Secondly, the use of KNIME nodes for data integration ensures that the data remains consistent and unaltered throughout the process. The automated nature of the nodes minimises the risk of human error during data transfer, ensuring that the integrity of the data is maintained. Furthermore, KNIME's flexibility in handling various data formats and its compatibility with a wide range of modelling tools make it an ideal platform for researchers working in nanoinformatics. Researchers can customise the nodes to suit their specific data types and modelling needs, which allows for a more tailored and effective analysis. Section 4.1 provides examples of KNIME nodes developed for nanoinformatics data and model integration, while WorldFAIR D4.1 provided an overview of the application of KNIME nodes for (nanomaterials) data FAIRification, which is being finalised as a journal publication.

3.6. Deployment of nanoinformatics models as FAIR software

3.6.1. Deployment as web applications

The development and implementation of web interfaces for all models mark significant advances in making complex computational tools more accessible and user-friendly. These web interfaces serve as essential gateways for researchers and practitioners to interact with various models, simplifying what would otherwise be intricate and often challenging computational processes. The primary advantage of these web interfaces lies in their ability to present complex data and models in a more comprehensible and visually engaging manner. They typically feature intuitive navigation systems that guide users through the functionalities of the models, regardless of the users' technical background. This aspect is particularly crucial in a field as diverse and specialised as nanoinformatics, where users range from experienced computational scientists to practitioners and researchers who may not have extensive programming skills. Interactive elements such as graphs and charts are integral components of these interfaces, offering dynamic visualisations of data and model outputs. Users can interact with these elements, such as by adjusting parameters or selecting specific data points, to gain deeper insights into the models' workings and results. This interactivity not only enhances the understanding of the models but also allows for a more hands-on approach to data analysis, making the process more engaging and informative. Input-output fields are another key feature of these web interfaces, allowing users to easily input data into the models and view the results. This simplification of data entry and retrieval is a significant improvement over more traditional, command-line driven interactions which can be daunting and error-prone for less experienced users. The ability to see real-time results as data are inputted and parameters are adjusted provides immediate feedback, which is invaluable for iterative processes of model tuning and hypothesis testing.

Web interfaces often include features such as data upload and download capabilities, model configuration options, and links to supplementary resources and documentation. Such comprehensive features ensure that users not only have easy access to the models but also have all the necessary tools and information at their disposal to effectively use these models. In these ways, the development of web interfaces for nanoinformatics models represent a significant stride towards democratising access to sophisticated computational tools in the field. By providing intuitive, interactive, and visually appealing platforms, these interfaces significantly enhance the user experience, making complex nanoinformatics models more approachable and usable for a broad spectrum of users. This accessibility is key to fostering broader adoption and application of nanoinformatics models across various research and application domains.

All web applications developed within the Enalos platform are developed using Java and the [zk framework](#)³⁸. An early example is the web application of the model for prediction of a protein (carbonic anhydrase, CA) binding to multi-walled carbon nanotubes (MWCNTs) and the toxicity of the resultant decorated-MWCNTs (Varsou et al., 2019), whereby the models are available for public

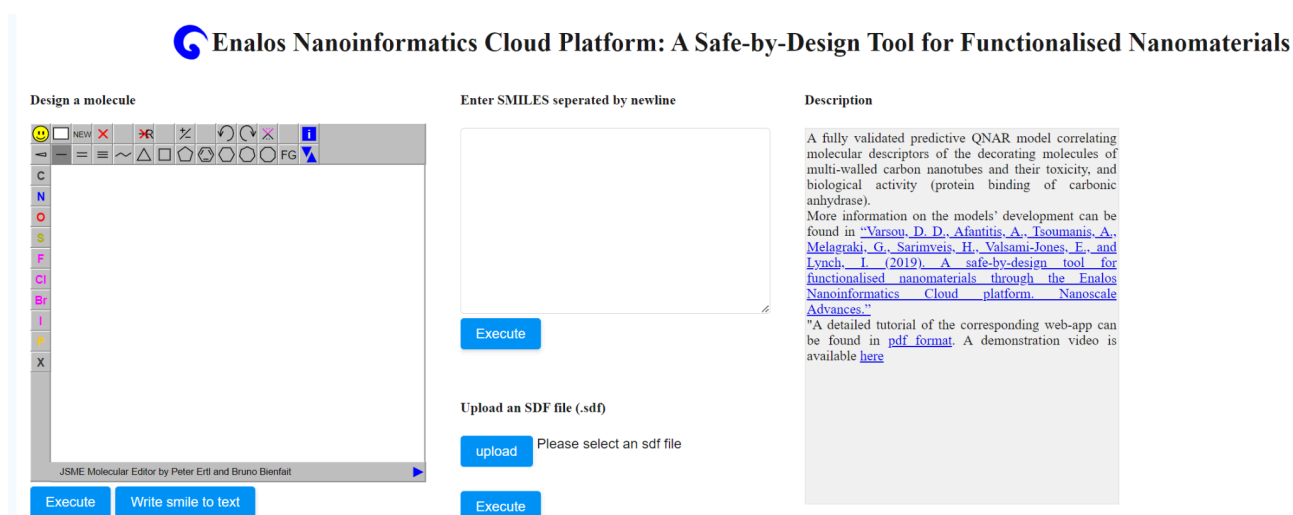


Figure 10: Enalos Nanoinformatics Cloud platform user-friendly interface for the web application of the models (<https://www.enaloscloud.novamechanics.com/EnalosWebApps/CNT/>). Users can simply draw the chemical structure of the decorating ligand, or upload a Structure-Data File³⁹ containing the molecular structure(s) of interest. The user can insert one or several structures of compounds being considered as potential decorating molecules for MWCNTs and get, within seconds, the prediction of the CA binding and their toxicity profile, along with a warning on the reliability of the predictions based on the models' domain of applicability limits. The user has three different options for providing the structures of the compounds to be screened: (i) by drawing the chemical structure of interest, (ii) by

³⁸ The ZK open source Java framework for enterprise web and mobile apps: <https://www.zkoss.org/>

³⁹ SDF connection table files refer to the Structure-Data File format defined and documented by Biovia as part of the MOL format family and used de facto in the community. <https://discover.3ds.com/ctfile-documentation-request-form>

entering the SMILES notation of the compounds in the appropriate field or (iii) by uploading an .sdf file with a batch of compounds. A tutorial is available at: <https://enaloscloud.novamechanics.com/EnalosWebApps/CNT/instructions.zul>.

use and verification through the Enalos Cloud platform⁴⁰ and can be used in order to observe the effects of the different inputs (decorating molecule structures, or ligands) on the prediction of CA binding to the MWCNTs and the toxicity of the resultant decorated-MWCNTs. The user-friendly web service, shown in Figure 10, facilitates the computer-aided design of novel MWCNTs by the interested users (computational experts or not); the Enalos Cloud platform can be easily accessed and directly explored by anyone interested in MWCNTs design to optimise functionality and safety (*i.e.* safe-by-design), without any need for prior programming skills.

We note that all of the options for providing chemical structures noted in Figure 10 have nuanced dependencies on the chemical structure notation, depending on implementation in the specific tools used that may not be indicated but which may have an impact for specific issues of conformation, stereoconfiguration, etc. An additional recommendation, provided by the WorldFAIR Chemistry case study (WP03)⁴¹ during review of the current document is that a validation and checking step, that includes cross-checking of the InChI for the uploaded chemical structures, should be implemented in the model to verify all incoming chemical structures and molecular notations.

By providing intuitive, interactive, and visually appealing platforms, these interfaces significantly enhance the user experience, making complex nanoinformatics models more approachable and usable for a broad spectrum of users. This accessibility is key to fostering broader adoption and application of nanoinformatics models across various research and application domains.

3.6.2. Deployment via APIs

The implementation of APIs for all models is a strategic advance that greatly augments the functionality and applicability of nanoinformatics models. APIs enable the efficient interaction between nanoinformatics models and a diverse array of software tools and platforms. This integration is essential in a field where multidisciplinary approaches and complex computational analyses are the norm.

APIs provide programmable access to the models, meaning that they allow users and developers to automate a range of tasks that would otherwise require manual intervention. This automation is key to streamlining processes such as data ingestion, model execution, and result analysis. By utilising APIs, repetitive and time-consuming tasks can be efficiently managed, allowing researchers and practitioners to focus on more critical aspects of their work such as interpreting

⁴⁰ <http://enalos.insilicotox.com/CNT/>

⁴¹ <https://worldfair-project.eu/chemistry-wp3/>

results and developing new hypotheses. Furthermore, APIs are instrumental in facilitating data exchange between different systems. In the area of nanoinformatics, where data from various sources often need to be collated and analysed, APIs provide a standardised method for data transfer. This standardisation ensures that data can be seamlessly moved between different models and tools, reducing the risk of data loss or corruption and ensuring the integrity of the data is maintained. Another significant advantage of APIs is their role in enabling the creation of complex workflows. Nanoinformatics research often involves intricate processes that require the integration of multiple models and tools. APIs allow these different components to communicate and work together harmoniously, enabling the development of sophisticated workflows that can handle complex analyses. This capability is particularly valuable for advanced research projects that require the integration of multiple data types and models to draw comprehensive conclusions.

The provision of well-documented and standardised APIs is thus crucial for making nanoinformatics models FAIR. Good documentation ensures that users understand how to effectively utilise the APIs, while standardisation guarantees compatibility across various platforms and tools. This compatibility is essential for ensuring that nanoinformatics models are interoperable, meaning they can be easily integrated and used in conjunction with other software tools. Such interoperability is essential for fostering collaborative research and facilitating the exchange of information and resources within the scientific community. The APIs enable innovation and efficiency. By allowing for automated task execution, seamless data exchange, and the creation of complex workflows, APIs significantly enhance the utility and applicability of nanoinformatics models. They facilitate a level of interoperability and flexibility that is crucial for advancing research in this multifaceted and rapidly evolving field.

Within NanoCommons, which aimed to integrate models and software from a range of different sources, we did not completely define an API specification but instead provided guidelines on how existing APIs can be improved to achieve alignment and interoperability between the services, by reviewing existing APIs from the project partners' tools and more general API concepts, finding similarities in these and building on top of them semantic annotations, which can be exploited when linking multiple services to understand the capacities of the service and the required input and generated output, as described in NanoCommons Deliverable D4.2 Initial [NanoCommons] APIs⁴². The Enalos and Isalos models developed as part of NanoCommons, NanoSolveIT and related projects generally use RESTful APIs, as described below.

The NanoCommons and NanoSolveIT projects offer a suite of RESTful APIs for various nanoinformatics models, each with a dedicated link. These APIs are tailored to the needs (inputs/outputs) of the specific models, meaning that there is a unique API for each model, allowing for targeted and specialised access to the functionalities of that particular model. Below are some examples of these APIs:

⁴² <https://zenodo.org/records/10473593>

NanoCommons APIs:

1. **NanoXtract Model RESTful API:** This API is designed for the NanoXtract model, which focuses on the extraction of nanoparticle features. Access it at [NanoXtract API](#).
2. **MS³bD Zeta Potential RESTful API:** This API is specific to the MS³bD model, which is used for calculating the zeta potential of nanoparticles. Access it at [MS³bD API](#).

NanoSolveIT APIs:

1. **Ecotox Models RESTful API:** This API pertains to the Ecotox models, which are used for ecological toxicity prediction. Access it at [Ecotox API](#).
2. **Facet Cytotoxicity RESTful API:** This API is for the Facet Cytotoxicity model, which focuses on cytotoxicity analysis. Access it at [Facet Cytotoxicity API](#).

The decision to utilise REST (Representational State Transfer) for the development of APIs in the context of nanoinformatics is grounded in its inherent qualities, including its simplicity and versatility, that align well with the needs of this field. REST APIs employ standard HTTP methods such as GET, POST, PUT, and DELETE, which are universally understood and widely used. This familiarity makes REST APIs relatively straightforward to understand and implement, contributing to their broad adoption. The use of these standard methods also ensures seamless integration with a diverse array of tools and platforms, making REST APIs versatile and adaptable to various technological environments. Another critical aspect of REST APIs is their statelessness. In a RESTful architecture, every request from a client to a server is self-contained, carrying all the necessary information to understand and process the request. This means that the server does not need to maintain any client state, thereby simplifying the server design and enhancing scalability and maintenance. This statelessness is particularly advantageous in scenarios where server resources need to be optimally managed and where the system's scalability is a priority. Furthermore, REST APIs offer significant flexibility in terms of data formats. While JSON (JavaScript Object Notation) is commonly used for data exchange due to its lightweight and easy-to-parse structure, REST APIs are not limited to JSON alone and can support other data formats, including XML (eXtensible Markup Language). This flexibility is particularly important in the field of nanoinformatics, where data can be highly diverse and complex. The ability to handle various data formats ensures that REST APIs can accommodate the specific data representation needs of different nanoinformatics models. Lastly, the alignment of REST with web standards is a compelling reason for its choice in nanoinformatics. RESTful APIs are inherently designed to work seamlessly with the web, making them an ideal choice for web-based applications and services. This is particularly beneficial for nanoinformatics models, many of which are accessed and utilised through web platforms. The web-centric design of REST APIs ensures that they integrate well with the online ecosystem, facilitating access and interaction with nanoinformatics models over the internet. The selection of REST for API development in nanoinformatics is driven by its simplicity, statelessness, flexibility in handling diverse data formats, and alignment with web standards, making it a fitting choice for the efficient and effective deployment of these advanced computational models.

3.7. Automated integration of model outputs back into nanoinformatics databases

As data undergo iterative modelling and receive feedback from various sources, they are refined and updated, thereby increasing their accuracy and relevance, resulting in a dynamic cycle where data are not static but constantly evolving. This process of continuous improvement not only enriches the individual datasets but also contributes to the growth of a comprehensive repository of high-quality data. Such a repository becomes an invaluable resource for researchers, offering a wealth of information that can be tapped into for a variety of nanoinformatics applications. Moreover, the updated and enriched datasets that result from this iterative process are better positioned to support advanced modelling techniques, including machine learning and predictive analytics. The enhanced quality and depth of data foster more accurate and reliable model outcomes, which in turn feed back into the cycle, further improving the datasets.

The FAIRification of nanoinformatics resources also plays a critical role in future-proofing the field. By adhering to these principles, researchers ensure that the data and tools they develop today will remain relevant and usable in the future, even as technologies and research paradigms evolve. This foresight is crucial in a field characterised by rapid technological advancements and changing research priorities.

The role of feedback mechanisms, particularly for predictions and validation statistics generated by modelling tools, is pivotal in shaping the continuous evolution and refinement of these models. Once the data has been processed through these tools, the resulting predictions and their corresponding validation statistics emerge as key indicators for assessing the model's accuracy and reliability. These metrics provide critical insights into how well the model performs and whether its predictions align with empirical data or established theoretical expectations. The automation of feeding these results back into databases, such as NanoPharos, plays a significant role in enhancing the lifecycle of the models used in nanoinformatics. By systematically updating these databases with the latest findings and predictive accuracies, along with any ancillary metadata generated during the modelling process, the models can be continually refined and improved. This iterative process of model validation and updating ensures that models remain relevant, accurate, and efficient in analysing and predicting nanomaterials' behaviours and properties.

Crucially, this feedback mechanism enhances the reusability of data. When new findings and validation statistics are integrated back into the database, they enrich the existing datasets. This enrichment does not just add volume to the data; it also enhances the depth and breadth of the information available, providing future users with a more robust and comprehensive dataset for their modelling exercises. For instance, new predictive accuracies can offer insights into specific conditions or parameters under which a model performs best, while additional metadata can provide context that may be crucial for interpreting the data correctly. Moreover, this process of updating the database with new information creates a dynamic data repository that evolves over time. It reflects the latest developments and understandings in the field of nanoinformatics,

making the database an ever-more valuable resource for researchers and practitioners. This dynamic nature of the database also fosters a collaborative scientific environment, where data and insights are continuously shared, contributing to the collective advancement of the field.

The feedback of predictions and validation statistics into databases is a process that transcends mere data updating. It embodies a proactive approach to model improvement, data enrichment, and knowledge sharing in nanoinformatics. This process ensures that the data and models used in this field remain cutting-edge, reliable, and increasingly useful for a wide array of research and practical applications in nanotechnology.

3.8. Sustainability of models and software via containerisation as microservices

The rapid pace of technological development means that it is essential to design software with sustainability and longevity in mind, and to enhance interoperability. The interoperability of nanoinformatics models is further enhanced through encapsulation of the models and their software into [Docker containers](#)⁴³ which can be easily integrated into larger systems and workflows. This is especially beneficial in microservices architectures, where different services (in this case, models) communicate with each other over a network. Making nanoinformatics models available as Docker containers with [OpenAPI specifications](#)⁴⁴ is a forward-thinking approach that addresses several key challenges in computational research. It also aligns well with the current trends in software engineering and data science, where there is an increasing emphasis on creating flexible, reusable, and reliable computational environments. However, while it significantly enhances portability, reproducibility, and interoperability, containerisation also brings forth considerations in sustainability and complex system integrations, necessitating ongoing efforts in optimisation and management. Maintaining and updating Docker containers to keep up with the evolving models and their environments can be resource-intensive. Another challenge is ensuring that the models remain compatible with various systems and other microservices over time.

3.8.1. Docker containers for model deployment

Docker³⁶, a popular containerisation platform, enables the encapsulation of models along with their entire runtime environment. This encapsulation includes the model itself, its dependencies, libraries, and any other necessary configurations. Packaging a model into a Docker container ensures that the model can be run consistently across different computing environments. This is because the container includes everything the model needs to run, isolated from the underlying system variations. The use of Docker containers enhances the portability of nanoinformatics models. Researchers and developers can easily share these containers, and users can run them on different machines without worrying about compatibility issues. This ease of transfer and deployment significantly accelerates the process of model testing and implementation.

⁴³ <https://www.docker.com/>

⁴⁴ <https://swagger.io/specification/>

3.8.2. Reproducibility and OpenAPI specification

Reproducibility is a critical aspect of scientific models, and Docker further strengthens this by ensuring that the computational environment is consistent every time the model is run. This consistency is vital for verifying and validating results in different scenarios and by different users. Complementing Docker, OpenAPI specifications provide a framework for describing the APIs of these models in a standard, language-agnostic way. This means that the model's interface can be clearly defined and understood, regardless of the programming language used to implement it. OpenAPI ensures that the model's functions are well-documented and can be interacted with consistently, which is crucial for integrating the model into various workflows and systems.

3.8.3. Example: Docker in practice

As a real example, the NanoSolveIT e-platform microservices have been packaged into [Docker containers](#), effectively encapsulating all necessary system dependencies. Adopting Docker ensured broad adaptability, extensive collaboration capabilities, and easy access to software packages through public repositories such as <https://hub.docker.com/>. NanoSolveIT deliverable report D7.5⁴⁵ describes how this technology not only facilitated rapid development of the NanoSolveIT cloud applications but also enabled the deployment of the services in standalone mode for users who prefer not to share their data to a Cloud. Additionally, the report highlighted the critical role of [Kubernetes](#)⁴⁶, the *de facto* platform for numerous open-source projects, as the container orchestration tool for the NanoSolveIT platform. Kubernetes further strengthened the capabilities of the standalone version. It offers unparalleled support and compatibility, allowing NanoSolveIT services to be deployed on any infrastructure running Kubernetes.

3.9. Maximising model utilisation through provision of training materials

The provision of comprehensive and step-by-step training materials to accompany all developed nanoinformatics models plays a crucial role in facilitating effective utilisation and understanding of these sophisticated tools. Within the NanoCommons and NanoSolveIT projects, the minimum documentation to accompany a developed model, tool or software was: i) a step-by-step user guidance document, which was ideally also accompanied by a demonstration video; and ii) a descriptive documentation of the principles underpinning the model, utilising the QSAR model report format (QMRf) for data-driven models and the emerging MODA standard for physics-based models (as described in Section 2). In most cases, a scientific publication also accompanied the models, as this was a means to announce the model and accompanying software to the scientific and stakeholder communities.

⁴⁵ NanoSolveIT D7.5: www.doi.org/10.5281/zenodo.10468032

⁴⁶ <https://kubernetes.io/>

Case studies are another valuable aspect of training materials, providing insights into how the models have been applied in real research scenarios, demonstrating their capabilities and potential applications. For experienced users, training materials can delve into advanced features and applications of the models. This might include guidance on customising the models for specific research needs, integrating the models with other tools and platforms, or using advanced computational techniques. By providing user guides, tutorials, case studies, and FAQs, these materials ensure that all users, regardless of their prior experience with the models, can access, understand, and utilise these complex tools to their fullest potential. This comprehensive support is essential for advancing research and collaboration in the field of nanoinformatics.

In general, model and/or software developers should provide this documentation and training materials, but there is potentially an opportunity for the user community to crowd-source some of these materials, or to supplement the official materials with practical user experience and feedback. Additionally, there is an opportunity to couple the training on use of the tools and software with the fundamental lessons of modelling and documentation presented herein. A recommendation for a next step is to develop these more in-depth training materials based on the current deliverable.

4. Examples of best practice from the nanosafety community

4.1. NanoCommons KNIME nodes

Within the NanoCommons research infrastructure project, a key goal was to develop best practice and streamlined workflows for integrating curated datasets into modelling tools and the resulting predictions back into the databases. One of the approaches chosen was to develop KNIME nodes for each model, tool and database, developed within the NanoCommons project, to allow linkage into user-friendly and fully automated workflows. As discussed above, KNIME is a graphical user interface (GUI)-based analytics platform aimed to create data analysis and modelling workflows combining different tools, without the need for user-provided coding.

The KNIME nodes developed in Nanocommons (Figure 11) include thirteen nodes that give access to the NanoCommons Knowledge Base to various models and to/from other databases, three nodes that are related to the estimation of the risk of triggering adverse outcome pathway (AOP) 173 (Lung Fibrosis) in mice due to exposure to 20 nm TiO₂ nanomaterials, two nodes that allow users to perform image analysis through the [NanoXtract](https://nanoxtract.cloud.nanosolveit.eu/)⁴⁷ image analysis tool for a single Transmission Electron Microscopy (TEM) image or a batch of TEM images, two nodes that give access to the predictive QNAR/kNN models correlating molecular descriptors of the decorating molecules of MWCNTs and their toxicity, and biological activity (see also Figure 10 which

⁴⁷ <https://nanoxtract.cloud.nanosolveit.eu/>

presented this model), and one node that gives access to the *k*NN/read-across model for the prediction of nanomaterials zeta-potential based on the nanomaterial core composition, the main elongation and the pH of the medium⁴⁸.

The KNIME nodes giving access to the NanoCommons Knowledge Base provide information about:

- The datasets included in the Knowledge Base,
- The organisations that contributed to the Knowledge Base,
- The nanomaterials in the Knowledge Base that are accompanied by TEM images,
- The nanomaterials in the Knowledge Base for which physicochemical and/or toxicity data are available,
- The available variables (toxicity and/or physicochemical).

The NanoCommons KNIME nodes can be easily linked with other KNIME nodes and can be used in data management, processing and modelling workflows. For instance, the models for the prediction of the bioactivity and toxicity of ligand surface-decorated MWCNTs (shown in Figure 10, above), were enwrapped as NanoCommons KNIME nodes. Thus, they can be directly used in predictive nanoinformatics workflows to support the design of safe nanomaterials - so-called safe by design or SbD. To initiate a prediction using the CNT nodes, the user must provide one or several structures of compounds being considered as potential decorating molecules (ligands) for MWCNTs and upon running the KNIME workflow gets, within seconds, the prediction of the CA binding and their toxicity profile, along with a warning on the reliability of the predictions based on the models' domain of applicability limits. For this purpose, two nodes are available allowing either the input of an SDF file or the input of a SMILES list to represent molecular structure, noting again the dependencies of these notations as highlighted in Section 3.6.1, and the recommendation to cross-check the generated structures using their InChIs (and NanoInChIs in due course once the standard is formally adopted and implemented). When structures are uploaded in either notation, a prediction and its reliability is generated within seconds (Figures 12-14).

⁴⁸ <https://doi.org/10.1002/sml.201906588>

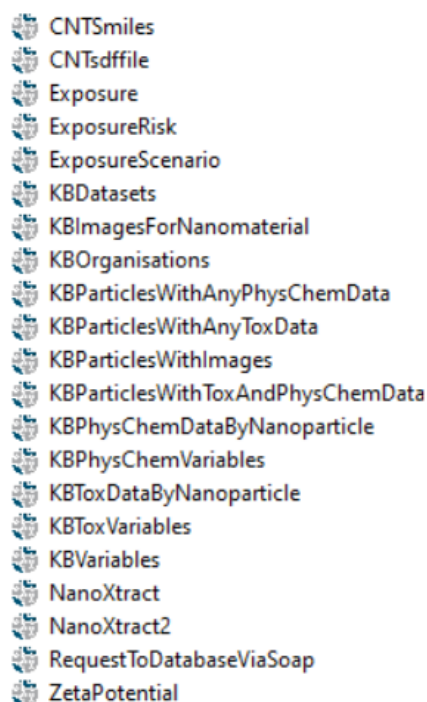


Figure 11: The NanoCommons KNIME nodes.

Row ID	column1
Row0	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row1	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(NC1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row2	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(Nc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row3	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(NCc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row4	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(N1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row5	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row6	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O</chem>
Row7	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(NCCCC)=O)N[H])=O</chem>
Row8	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(N(CCCC)CCCC)=O)N[H])=O</chem>
Row9	<chem>c1cccc(c1)C(Oc1cccc(cc1)CC(C(NC1CCCC1)=O)N[H])=O</chem>

Figure 12: Example of input data to the NanoCommons CNTSmiles KNIME node utilising [SMILES notation](https://nepis.epa.gov/Exe/ZyPDF.cgi/2000CAUR.PDF?Dockey=2000CAUR.PDF)⁴⁹ to collect the information about each CNT.

⁴⁹ <https://nepis.epa.gov/Exe/ZyPDF.cgi/2000CAUR.PDF?Dockey=2000CAUR.PDF>

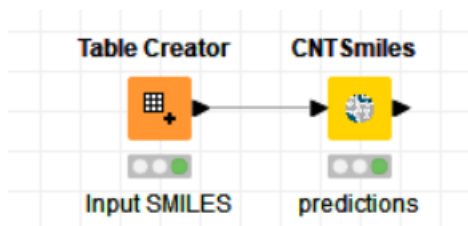


Figure 13: Use of the NanoCommons CNTSmiles KNIME node. The green dots indicate that the node has run correctly and made the predictions, which are shown in Figure 14.

Row ID	S id	S activity	S activityReliability	S toxicity	S toxicityReliability
Row0	Row0	non-binder	reliable	toxic	unreliable
Row1	Row1	binder	reliable	toxic	reliable
Row2	Row2	binder	reliable	toxic	reliable
Row3	Row3	binder	reliable	toxic	reliable
Row4	Row4	binder	reliable	toxic	reliable
Row5	Row5	non-binder	reliable	toxic	unreliable
Row6	Row6	binder	reliable	toxic	unreliable
Row7	Row7	non-binder	reliable	non-toxic	reliable
Row8	Row8	binder	reliable	toxic	reliable
Row9	Row9	binder	reliable	toxic	reliable

Figure 14: Output results of the NanoCommons CNTSmiles KNIME node showing the predictions for each CNT and whether the prediction is reliable (within the domain of applicability) or not.

4.2. NanoCommons User Handbook

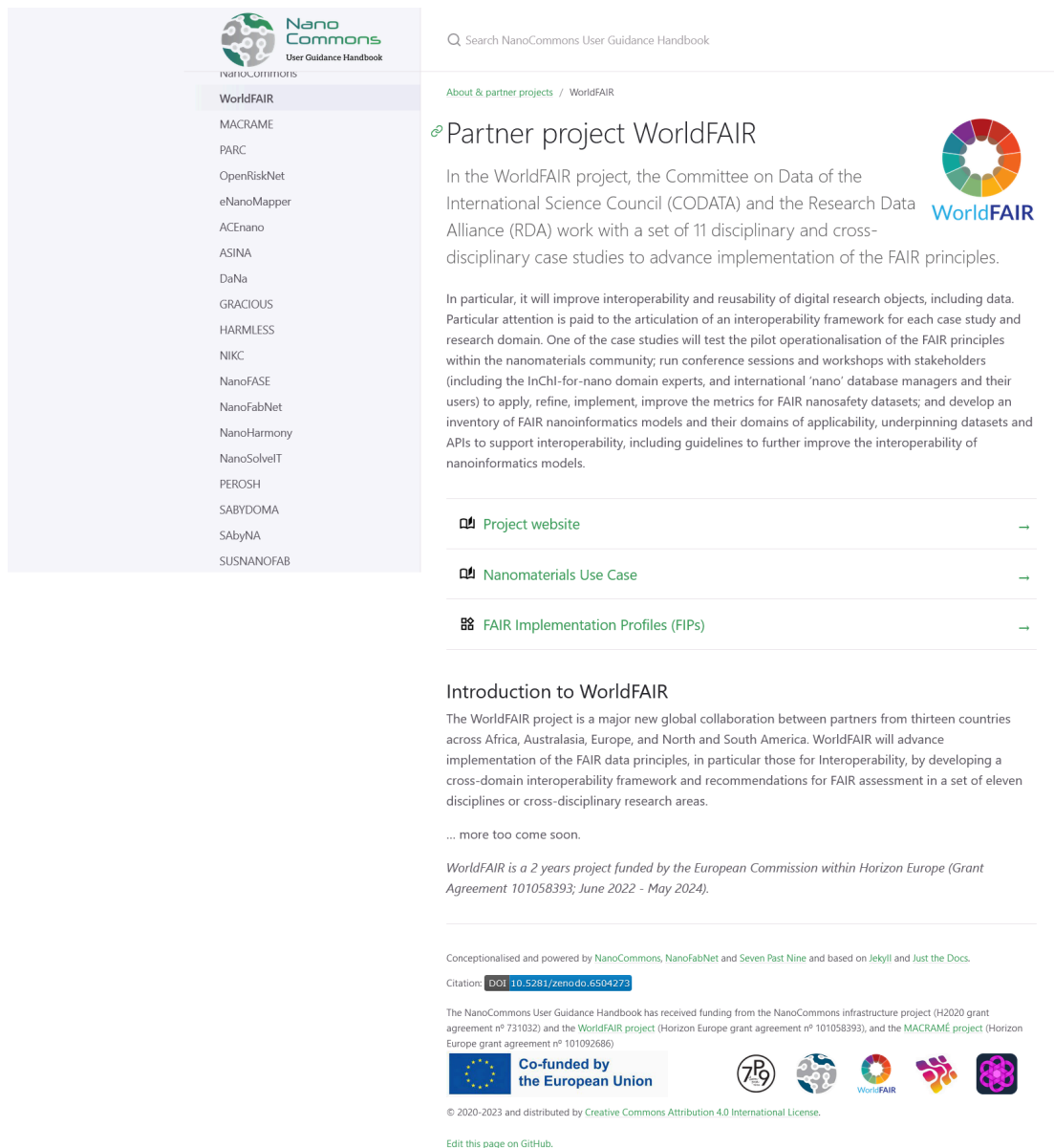
The NanoCommons [User Handbook](#) was conceived as a self-training platform for collecting and organising knowledge and training materials for nanomaterials specific data management and nanoinformatics methods. The handbook started by listing and describing the individual services and grouping them into application areas and in that way, providing guidance for users of all experience levels by starting with simpler, easier to learn tools or applications of these tools and then continuing through to increasingly complex situations or applications of the tools and services. However, it was then recognised that the handbook will have more value and will become a useful resource in many more situations if, in addition to the information on how to access the tools and other training materials (such as manuals and tutorials), information and guidance on the underlying theory, concepts and implementations are also provided as context and supporting knowledge.

This led to the first set of topics and the layout of the User Handbook landing page (see Figure 14), including [Data management](#), [FAIRification](#), [Nanoinformatics](#), [Training courses](#), and [Demonstration](#)

[cases](#), which has, during the course of the WorldFAIR project been further updated, to include sections on [Risk Assessment](#), [Nano \(risk\) Governance](#), [Safe and Sustainable by Design](#), [NanoFabrication](#) and information on relevant (partner) [projects](#) who have contributed to the development and provided content for the handbook. Figure 15 shows the example of the WorldFAIR page, and acknowledges that onward development since the end of NanoCommons in June 2022 has been supported by WorldFAIR.

To show how this information is presented and what type of material can be added, the ‘Data management’ section is described in more detail and snapshots from the current version are presented. In the User handbook, methods and processes are described in context with the available tools, coming both from within and beyond NanoCommons. The aim is to enable users to apply the described process, ideally supported by the described tool to improve the (re)use of nanomaterials safety-related data and improve computational analysis and predictive modelling approaches for integrated risk assessment and application of the Safe-by-Design framework⁵⁰. This is done with a general introduction to the area, which in the case of data management is a replication of the data management lifecycle (see Figure 17, below).

⁵⁰ [JRC Publications Repository - Safe and sustainable by design chemicals and materials - Framework for the definition of criteria and evaluation procedure for chemicals and materials \(europa.eu\)](#)



The screenshot shows the NanoCommons User Guidance Handbook interface. On the left is a sidebar with a list of projects: WorldFAIR (highlighted), MACRAME, PARC, OpenRiskNet, eNanoMapper, ACEnano, ASINA, DaNa, GRACIOUS, HARMLESS, NIKC, NanoFASE, NanoFabNet, NanoHarmony, NanoSolveIT, PEROSH, SABYDOMA, SAbYNA, and SUSNANOFAB. The main content area is titled 'Partner project WorldFAIR' and includes a search bar, a description of the project, and links to the project website, Nanomaterials Use Case, and FAIR Implementation Profiles (FIPs). The footer contains funding information, a citation, and logos for the European Union, 7P, and WorldFAIR.

WorldFAIR

MACRAME

PARC

OpenRiskNet

eNanoMapper

ACEnano

ASINA

DaNa

GRACIOUS

HARMLESS

NIKC

NanoFASE

NanoFabNet

NanoHarmony

NanoSolveIT

PEROSH

SABYDOMA

SAbYNA

SUSNANOFAB

Search NanoCommons User Guidance Handbook

About & partner projects / WorldFAIR

Partner project WorldFAIR

In the WorldFAIR project, the Committee on Data of the International Science Council (CODATA) and the Research Data Alliance (RDA) work with a set of 11 disciplinary and cross-disciplinary case studies to advance implementation of the FAIR principles.

In particular, it will improve interoperability and reusability of digital research objects, including data. Particular attention is paid to the articulation of an interoperability framework for each case study and research domain. One of the case studies will test the pilot operationalisation of the FAIR principles within the nanomaterials community; run conference sessions and workshops with stakeholders (including the InChI-for-nano domain experts, and international 'nano' database managers and their users) to apply, refine, implement, improve the metrics for FAIR nanosafety datasets; and develop an inventory of FAIR nanoinformatics models and their domains of applicability, underpinning datasets and APIs to support interoperability, including guidelines to further improve the interoperability of nanoinformatics models.

- [Project website](#)
- [Nanomaterials Use Case](#)
- [FAIR Implementation Profiles \(FIPs\)](#)

Introduction to WorldFAIR

The WorldFAIR project is a major new global collaboration between partners from thirteen countries across Africa, Australasia, Europe, and North and South America. WorldFAIR will advance implementation of the FAIR data principles, in particular those for Interoperability, by developing a cross-domain interoperability framework and recommendations for FAIR assessment in a set of eleven disciplines or cross-disciplinary research areas.

... more too come soon.

WorldFAIR is a 2 years project funded by the European Commission within Horizon Europe (Grant Agreement 101058393; June 2022 - May 2024).

Conceptionalised and powered by NanoCommons, NanoFabNet and Seven Past Nine and based on Jekyll and Just the Docs.

Citation: DOI: 10.5281/zenodo.6504273

The NanoCommons User Guidance Handbook has received funding from the NanoCommons infrastructure project (H2020 grant agreement n° 731032) and the WorldFAIR project (Horizon Europe grant agreement n° 101058393), and the MACRAME project (Horizon Europe grant agreement n° 101092686)

Co-funded by the European Union

© 2020-2023 and distributed by Creative Commons Attribution 4.0 International License.

[Edit this page on GitHub.](#)

Figure 15. Start screen of the User Handbook showing the WorldFAIR project landing page at <https://nanocommons.github.io/user-handbook/about-and-partner-projects/WorldFAIR>, with links to the project website, the WP04 case study and the nanomaterials FIP for commenting. This is currently being updated with further content including the NanoInChI progress, recordings related to the Nanomaterials FIP, the case study deliverables / milestones and publications to date, and more.



Figure 16. Introduction to the data management section of the NanoCommons User Handbook showing the data management lifecycle and a link to the original NanoCommons publication describing this (Papadiamantis et al., 2020).

After the brief introduction to each section, more specific topics are covered, integrating additional resources. To directly show what type of resources the user will be linked to, specific icons are utilised consistently throughout the User Handbook, including symbols to indicate web resources, videos, further text and links to external websites, as shown in Figure 17.







Icon	Content type
	Web application, software tool or other digital service
	Video
	Slides from presentations or posters
	Text documents like reports, papers or blog posts
	Extended material from external websites
	Any other type of content

Figure 17. Types of major resources integrated into the User Handbook. ‘Application’ means a specific tool or service, generally delivered via a web-based user-friendly interface.

4.3. Deployment of models / software to EOSC

To help guarantee sustainability of the NanoCommons Knowledge Infrastructure beyond its current funding lifetime, and consolidate its integration into the European infrastructure landscape, including the European Open Science Cloud (EOSC), a project was undertaken to develop a NanoCommons EOSC deployment. To achieve this, a strategic collaboration was formed with the H2020 [EGI-ACE project](#)⁵¹ to provide cloud resources for the hosting of the NanoCommons services on EOSC. EGI-ACE is providing computer resources to the EOSC Compute Platform and is contributing to the EOSC Data Commons through a federation of cloud compute and storage facilities, Platform as a service (PaaS) services and data spaces with analytics tools and federated access services. A summary of the process is presented here, as a basis for further development.

4.3.1. Approach to deployment

4.3.1.1. Cluster setup

⁵¹ EGI-ACE is funded by H2020 (2020-2023) with the objective to implement the Compute Platform of EOSC and contribute to the EOSC Data Commons by delivering an integrated solution that is aligned with major European cloud federation projects and high performance computing initiatives.

The VO setup was accompanied by a VO service level agreement and an operational level agreement. Figure 18 shows the Initial Cluster Configuration established for the NanoCommons EOSC deployment.

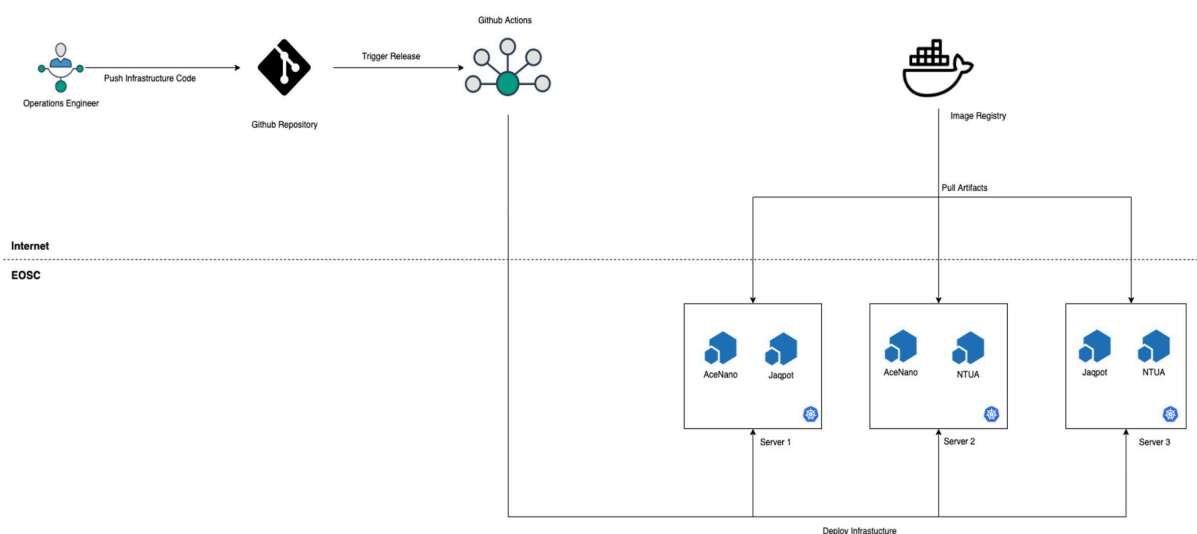


Figure 18. Initial Cluster Configuration established for the NanoCommons EOSC deployment.

4.3.1.2. Kubernetes environment

We established a Kubernetes environment to support the deployment and orchestration of NanoCommons resources on the cluster (architecture overview provided in Figure 19). The Kubernetes environment consists of four instances (one master node and three worker nodes) deployed on OpenStack. An ingress machine is deployed in front to provide additional proxying and a layer of security (running nginx natively).

⁵² <https://www.egi.eu/>

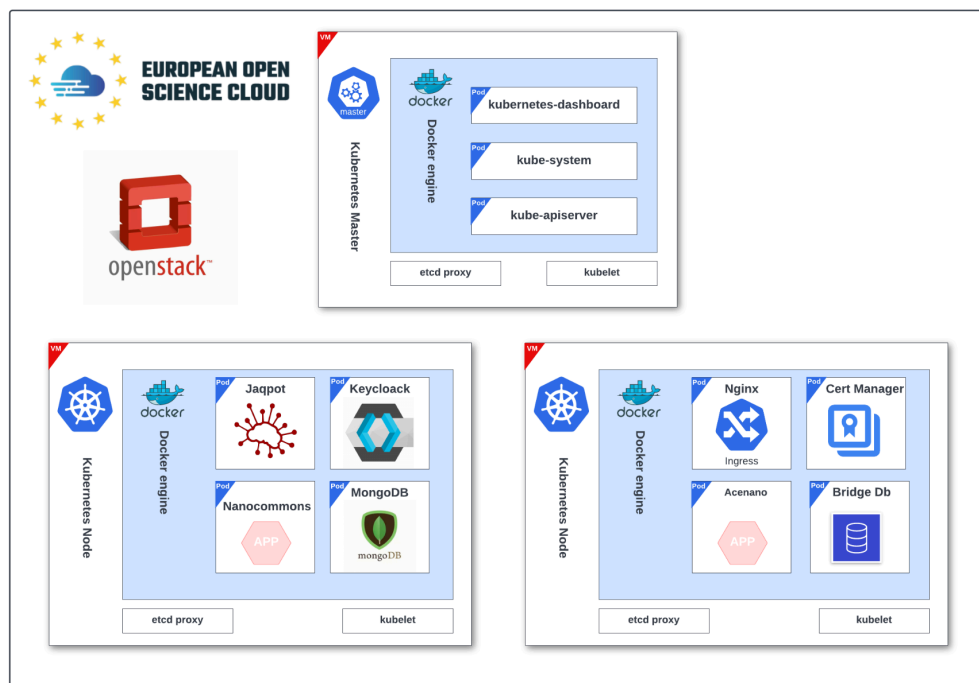


Figure 19. High level architecture for orchestration and deployment of NanoCommons resources to EOSC.

Deployments are tracked and versioned in a GitHub repository. YAML⁵³ files are used for configuration and deployment of Kubernetes resources. A future step forward could be shifting to Helm⁵⁴ which would provide additional possibilities such as templating and additional customisation.

The NanoCommons EOSC cluster was established with five initial resources deployed as follows:

- Jaqpot (NTUA) - <https://app.eosc.jaqpot.org/>
- NovaMechanics resources - five resources have been deployed to EOSC.
<http://www.enalosccloud.novamechanics.com/nanocommons.html>
- NanoCommons Knowledge Base (Biomax) -
https://ssl.biomax.de/nanocommons/cgi/login_bioxm_portal.cgi
- AOPwiki SPARQL (UM) -- <https://github.com/marvinm2/AOPWikiRDF>
- ACEnano (EwC) - <https://acenano.opentox.net/> established as a test deployment under the NanoCommons EOSC cluster.

⁵³ <https://yaml.org/>

⁵⁴ <https://helm.sh/>

4.4. The NanoSolveIT cloud platform

The NanoSolveIT cloud platform (<https://cloud.nanosolveit.eu>) is the tangible outcome of dedicated collaborative efforts and research conducted within the NanoSolveIT project. The cloud platform serves as a comprehensive hub for nanoinformatics and a means to democratise access to nanoinformatics by provision of user-friendly interfaces that require no programming skills from the user. The platform incorporates a wide range of computational modelling tools and functionalities that were specifically designed during the NanoSolveIT project. It provides users with a seamless and integrated environment to conduct exposure, hazard, and risk assessments. By combining state-of-the-art technologies, innovative *in silico* methods and advanced models, the platform empowers users to explore and understand the intricate characteristics and potential adverse effects of NMs on both human health and the environment. Furthermore, the NanoSolveIT cloud platform goes beyond individual modelling tools by offering a unified framework that streamlines the input of data and generation of predictions in response to the users' queries. It supports the implementation of nanoinformatics-driven integrated approaches, enabling users to assess the risks associated with NM exposure.

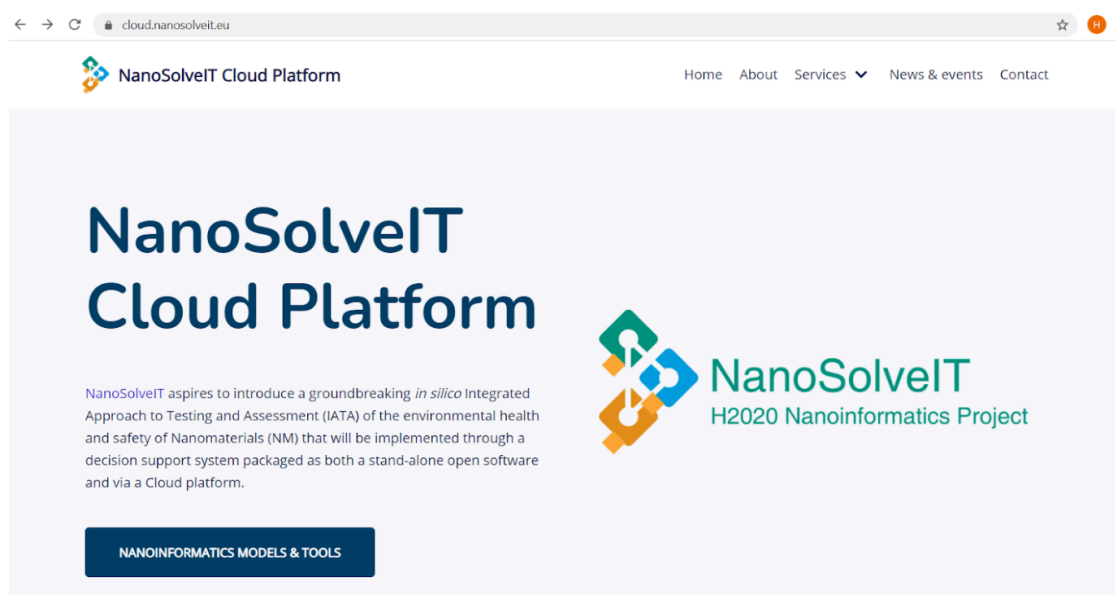


Figure 20: The landing page of the NanoSolveIT cloud platform.

The landing page of the NanoSolveIT Cloud Platform provides basic information about the project (Figure 20) and contains a summary of the various tools that are offered through the platform, which are grouped into seven categories (end-points or model types) (Figure 21 and Table 3). By clicking on “Learn More”, a new page appears which presents a generic description of this category,

describes all the services which belong to the category and provides links to the services that are hosted in the platform under this category.

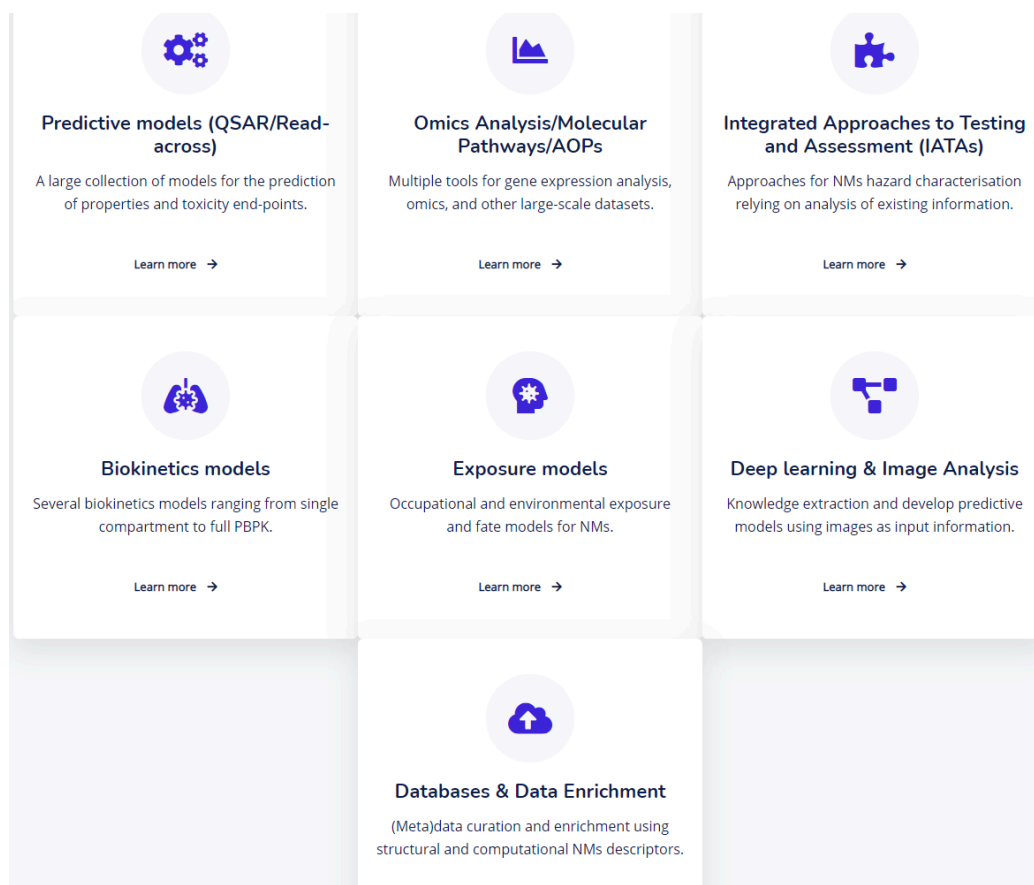


Figure 21. A screenshot of the NanoSolveIT cloud platform where the categories of services are presented. Currently there are seven categories covering the different types of models or services.

Table 3. Categories of nanoinformatics tools and services (currently) offered through the NanoSolveIT cloud platform.

1. Exposure Models	A number of occupational and environmental exposure and fate models for ENMs are offered through the NanoSolveIT platform with graphical user interfaces (GUIs) where users can define and simulate various occupational and environmental exposure scenarios.
2. Predictive Models (QSAR/Read-across)	A large collection of models for the prediction of NM properties and toxicity end-points are provided as ready-to-use web applications through the NanoSolveIT platform. The models have been developed using various approaches, including machine learning methods, computational read-across techniques and prediction networks (e.g., Bayesian networks).
3. Biokinetics models	Several biokinetics models ranging from single compartment to full physiologically based pharmacokinetic (PBPK) models for both humans and environmental species are provided with GUIs. Biokinetics models are used for the estimation of internal exposure and biologically effective doses for NMs.
4. Deep Learning & Image Analysis Tools	Deep Learning is a powerful machine learning technology that can extract knowledge and create predictive models using images as input information. Based on this technology, the NanoSolveIT consortium has developed computational workflows for the prediction of adverse effects due to exposure to NMs, which are offered through the platform along with other tools for image analysis and feature extraction as a means to enrich and gap-fill datasets for modelling and prediction.
5. Omics Analysis / Molecular Pathways / AOPs	The NanoSolveIT platform offers multiple tools for analysis of gene expression, transcriptomics, and other large-scale datasets, as well as access to molecular pathway and adverse outcome pathway (AOP) databases.
6. Databases & Data Enrichment	NanoSolveIT offers a complete solution regarding the maximisation of data's exploitation potential. The available databases and Knowledge Bases provide users the ability to upload, semantically annotate and enrich their data using established ontologies and a set of structural, molecular and atomistic descriptors. Users can get ready-for-modelling datasets to import into their computational workflows and/or link the data with online tools to perform automated analysis and results retrieval.
7. Integrated Approaches to Testing and Assessment (IATAs)	IATA are approaches for NM hazard characterisation that rely on an integrated analysis of existing information coupled with the generation of new information using testing strategies. We have developed workflows that combine and integrate external exposure models with biokinetics models for extracting the maximal amount of information among disparate models and to improve the source-to-exposure-to-dose estimate.

As described in Section 3.9 of this deliverable, detailed tutorials / training materials for the web tools and services offered within the NanoSolveIT cloud platform are available. The tutorials

provide in-depth, step-by-step instructions for utilising each individual web tool. The primary purpose of these tutorials is to aid users in acquiring a thorough comprehension of the features, functionality, and practical implementation of these tools. Through detailed explanations and illustrative examples, the tutorials aim to equip users with the essential knowledge and skills required to maximise the utilisation of the available web tools.

Each tutorial begins with a brief description of the tool or service and is then accompanied by a comprehensive walkthrough of the process. Visual aids, such as illustrations or diagrams, are included to assist users in following the instructions effectively and to successfully navigate and utilise the web tools offered by NanoSolveIT. It is important to note that each tutorial is self-contained and follows its own numbering system for tables and figures. This ensures clarity and consistency within each tutorial. By following these tutorials, users can learn the features, functionality, and best practices for using specific tools. The primary goal is to help users enhance their understanding and proficiency in utilising the NanoSolveIT cloud platform and thus to make the most out of the platform's nanoinformatics services. By accessing these tutorials, users gain a deeper understanding of the capabilities and potential applications of the NanoSolveIT platform, enabling them to leverage the platform effectively and facilitating their research and decision-making processes.

4.5. Integrated workflow - from model development to FAIR software

As part of our WorldFAIR WP04 case study on nanomaterials, we are currently finalising the implementation of a complete workflow for data curation and documentation, model development and documentation, deployment of the model as software via a web interface and API (to enhance accessibility and ensure interoperability) and its registration in the nanoinformatics model registry (see Section 3.3 above) to demonstrate the value of FAIRification of research software in the nanoinformatics domain. The steps involved are shown schematically in Figure 21, and map also to the detailed model development and validation SOP developed in NanoSolveIT and enriched via WorldFAIR WP04 (see Figure 7) to include the FAIRification steps for the data underlying the model, the model metadata and the software metadata and data.

The modelling approach combines different state-of-the-art computational methodologies including an automatic machine learning scheme that employs dose-response toxicity data for Ag, TiO₂, and CuO nanomaterials, and physics-based simulation on fully computational nanomaterials as a means to enrich the dataset with atomistic descriptors to capture their underlying structural properties. Furthermore, to overcome the limitations of data availability, we employed synthetic data generation techniques to expand the dataset and improve the representation of different nanomaterials classes, as a means to overcome data imbalance using reliable oversampling techniques that generate synthetic data samples, ensuring an equal representation of the classes in the dataset. Finally, the users' confidence in the generated results is ensured through a three-step

applicability domain approach that assesses the level of reliability of the prediction, i.e., utilising first the bounding-box and the leverage approach and then applying a local similarity approach proposed for the first time herein. According to this strategy, the reliability of each prediction can be classified as “good”, “moderate”, or “poor” based on a weighted scheme that incorporates the results of the three above-mentioned strategies. Finally, to strengthen the confidence of the interested end-users to the proposed model, both the data and the model are made freely available through the NanoPharos database and the Enalos Cloud platform, respectively, to be explored and used by the stakeholders, and are fully documented including their FAIR metrics score.

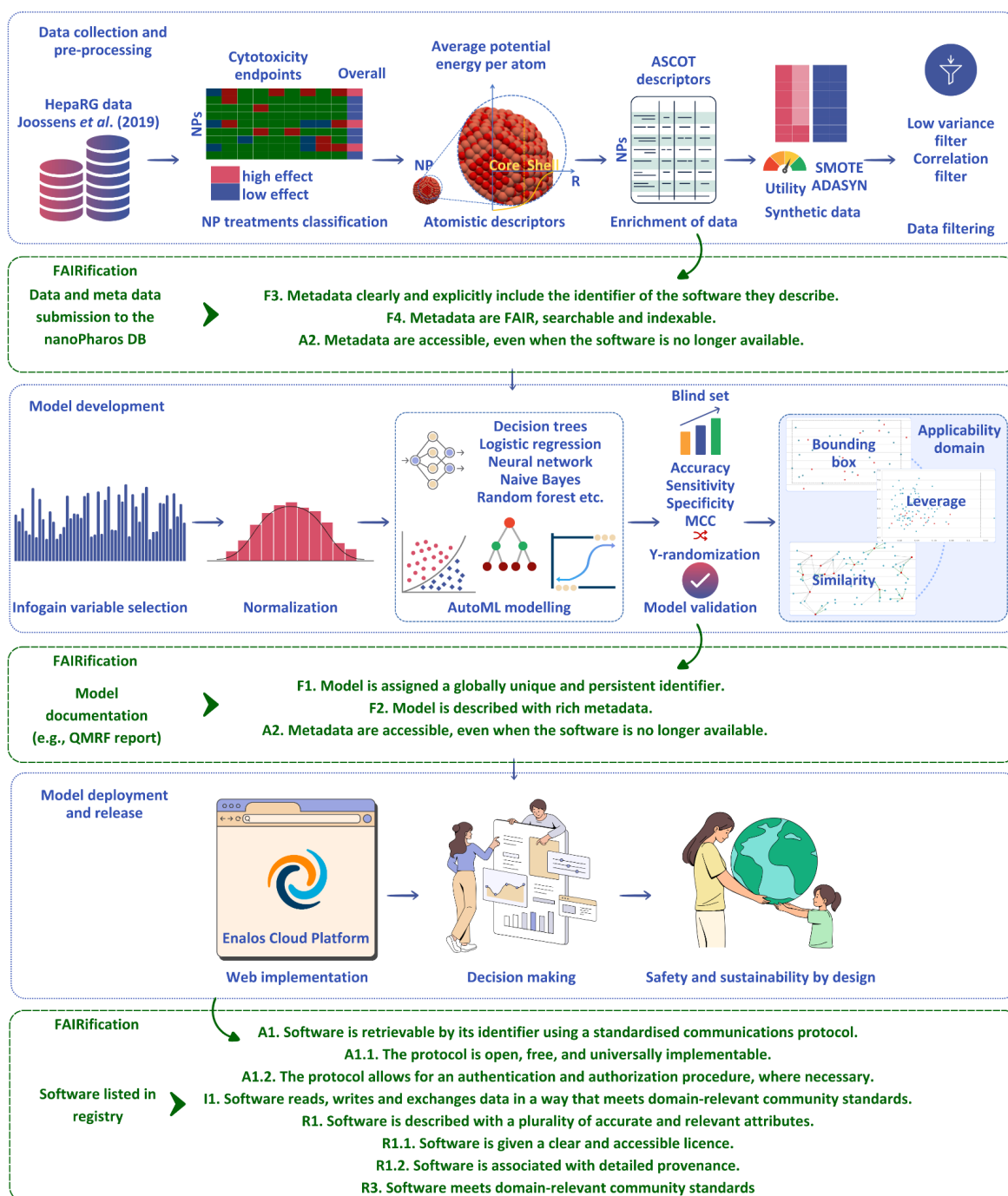


Figure 21: Schematic workflow of the data analysis, modelling of the nanomaterials toxicity endpoints, and release of the final model. Image created via canva.com. Adapted from Varsou *et al.*, 2024, Submitted.

5. Conclusions and recommendations

This deliverable report focuses on practical approaches to be implemented primarily by nanoinformatics models and software developers, to enhance the FAIRness of nanoinformatics tools and software and their underlying datasets. It also provides concrete examples of the suggested approaches that have already been implemented for nanomaterials safety assessment and nanoinformatics, and provides recommendations for next steps to further enhance the FAIRness of nanoinformatics models. The deliverable highlights the need for a structured, efficient, and collaborative approach to data management and usage in the field to support nanoinformatics development and acceptance for industrial and regulatory use. It underscores the importance of creating a research ecosystem where data is not just a static entity but a dynamic resource that grows and improves over time, thus driving innovation and discovery.

The importance of metadata for nanoinformatics models and software is highlighted, and the role of existing and emerging harmonised model documentation formats, including QMRF for QSAR and machine learning (data-driven) models and MODA for physics-based models, is presented. Work is underway within WorldFAIR to develop a prototype webtool for completion of MODAs that reduces the need for free-text which is error-prone and can hamper integration. Such harmonised documentation enriches the user's understanding of these models and their application, and will drive uptake by industrial stakeholders and facilitate regulatory validation and acceptance. These templates provide a detailed narrative of the model's purpose, design, evolution, and usage, which are crucial also for re-use of the models.

The use of KNIME in nanoinformatics represents a significant advancement in the field. It facilitates a more efficient and accurate integration of datasets into modelling tools, ensuring a smoother transition from data curation to application. This not only enhances the overall process of data analysis but also allows researchers to focus more on the interpretation and application of their findings, rather than on the mechanics of data preparation and integration.

Based on the information collated and summarised in this report, the following recommendations are provided for the nanoinformatics community and its stakeholders including industry and regulators.

Policy recommendations to support FAIRification of nanoinformatics:

- Documentation of the datasets underpinning nanoinformatics models is critical, and thus the MODA pre-standard for documenting physics-based models should be adopted as an equivalent to the QMRF for QSAR models, and the PBPK reporting format for Physiologically based pharmacokinetic (PBPK) models.
- Adoption of MODA as having equivalent importance to QMRF, will drive development of tools for automating the completion and validation of MODAs for wider range of models, and

support increasing harmonisation of the minimum reporting guidelines and metadata that accompanies nanoinformatics models, thus further embedding its adoption and regulatory acceptance.

- The QSAR model toolbox and other platforms supporting the deployment of QSAR models for regulatory purposes such work to implement the recommendations of the RDA FAIR4RS principles, and the additional considerations presented here for nanoinformatics models.
- Deployment of nanoinformatics models to the EOSC platform is possible and should be encouraged, but it is important to note that prior to regulatory acceptance of nanoinformatics models, the number of users is likely to remain below the critical mass needed to sustain a cluster of activity on EOSC. Additionally, extensive training materials targeting both nanoinformatics model developers and potential users would be needed, as the processes for engagement with EOSC are not easily accessible to non-experts / non-programmers.

Community action recommendations to support FAIRification of nanoinformatics:

- The nanoinformatics community needs to move beyond ‘black-box’ approaches and be able to demonstrate causality in model predictions, and chemical/particle-based descriptions of the model outcome (prediction). Consideration of quantitative *substance-activity* relationships rather than quantitative *structure-activity* relationships may provide a route towards this. Similarly, multi-criteria optimisation approaches can help to pinpoint the key nanomaterials properties that drive their toxicological effects.
- Nanoinformatics model developers are encouraged to apply the following steps to maximise the FAIRness of their models and software in the following ways:
 - Document the model and software utilising MODA or QMRF as rich metadata linked to existing ontology terms as well as documenting the underpinning datasets and their completeness and quality evaluation.
 - Register the model in the emerging Materials Models Registry being developed in WorldFAIR including its provenance and re-use / extension conditions. Registration will provide the model / software with a unique and persistent identifier.
 - Ensure access to the model / software through provision of a web interface, an application programming interface or a KNIME node, and/or provide a container and deploy it to EOSC or another cloud platform.
 - Provide user guides, tutorials and other training materials suitable for all users, regardless of their prior experience or technical competence, to enable the widest possible uptake of the models.

- Nanoinformatics models that utilise chemical structures, whether by drawing the chemical structure of interest, by entering the SMILES notation of the compounds, or by uploading an .sdf file should be aware that the conformation or stereoconfiguration might vary slightly, and thus automated checking and validation steps to confirm the chemical structures should be implemented, including cross-checking of the InChI for the uploaded chemical structures.
- There is an opportunity to couple the training on use of nanoinformatics tools and software with the fundamental lessons of modelling and documentation presented herein. Dedicated training resources for nanoinformatics models and software FAIRification should be developed for inclusion in Masters and PhD level training of (nano)informaticians.

6. Bibliography

- Afantitis A, Melagraki G, Isigonis P, Tsoumanis A, Varsou DD, Valsami-Jones E, Papadiamantis A, Ellis LJA, Sarimveis H, Doganis P, Karatzas P, Tsiros P, Liampa I, Lobaskin V, Greco D, Serra A, Kinaret PAS, Saarimäki LA, Grafström R, Kohonen P, Nymark P, Willighagen E, Puzyn T, Rybinska-Fryca A, Lyubartsev A, Jensen KA, Brandenburg JG, Lofts S, Svendsen C, Harrison S, Maier D, Tamm K, Jänes J, Sikk L, Dusinska M, Longhin E, Rundén-Pran E, Mariussen E, El Yamani N, Unger W, Radnik J, Tropsha A, Cohen Y, Leszczynski J, Hendren CO, Wiesner M, Winkler D, Suzuki N, Yoon TH, Choi J-S, Sanabria N, Gulumian M, Lynch I. NanoSolveIT Project: Driving nanoinformatics research to develop innovative and integrated tools for in silico nanosafety assessment. *Computational and Structural Biotechnology Journal*, 2020, 18, 583-602. <https://doi.org/10.1016/j.csbj.2020.02.023>
- Barker, M., Chue Hong, N.P., Katz, D.S. *et al.* Introducing the FAIR Principles for research software. *Sci Data* 9, 622 (2022). <https://doi.org/10.1038/s41597-022-01710-x>
- Basei G, Rauscher H, Jeliaskova N, Hristozov D. A methodology for the automatic evaluation of data quality and completeness of nanomaterials for risk assessment purposes, *Nanotoxicology*, 2022, 16:2, 195-216. <https://doi.org/10.1080/17435390.2022.2065222>
- Drobne D, Ciornii D, Hodoroaba VD, Bohmer N, Novak S, Kranjc E, Kononenko V, Reuther R. Knowledge, Information, and Data Readiness Levels (KaRLs) for Risk Assessment, Communication, and Governance of Nano-, New, and Other Advanced Materials. *Global Challenges*, 2023, 7, 2200211. <https://doi.org/10.1002/gch2.202200211>
- Dumit VI, Ammar A, Bakker MI, Bañares MA, Bossa C, Costa A, Cowie H, Drobne D, Exner TE, Farcas L, Friedrichs S, Furxhi I, Grafström R, Haase A, Himly M, Jeliaskova N, Lynch I, Maier D, Noorlander CW, Shin HK, Soler-Illia GJAA, Suarez-Merino B, Willighagen E, Nymark P. From principles to reality. FAIR implementation in the nanosafety community. *Nano Today*, 2023, 51, 101923. <https://doi.org/10.1016/j.nantod.2023.101923>
- Elberskirch L, Binder K, Riefler N. *et al.* Digital research data: from analysis of existing standards to a scientific foundation for a modular metadata schema in nanosafety. *Part Fibre Toxicol.* 2022, 19, 1. <https://doi.org/10.1186/s12989-021-00442-x>
- European Chemicals Agency, New Approach Methodologies in Regulatory Science. Proceedings of a scientific workshop, Helsinki, 19–20 April 2016. https://echa.europa.eu/documents/10162/21838212/scientific_ws_proceedings_en.pdf/a2087434-0407-4705-9057-95d9c2c2cc57

- European Commission, Directorate-General for Research and Innovation, Baas, A., *What makes a material function? – Let me compute the ways – Modelling in H2020 LEIT-NMBP programme materials and nanotechnology projects – Sixth version*, Baas, A.(editor), Publications Office of the European Union, 2017, <https://data.europa.eu/doi/10.2777/417118>
- Exner TE, Papadiamantis AG, Melagraki G, Amos JD, Bossa N, Gakis GP, Charitidis CA, Cornelis G, Costa AL, Doganis P, Farcas L, Friedrichs S, Furxhi I, Klaessig FC, Lobaskin V, Maier D, Rumble J, Sarimveis H, Suarez-Merino B, Vázquez S, Wiesner MR, Afantitis A, Lynch I. Metadata stewardship in nanosafety research: learning from the past, preparing for an “on-the-fly” FAIR future. *Front. Phys.* 2023, 11:1233879. <https://doi.org/10.3389/fphy.2023.1233879>
- Fernández-Cruz ML, Hernández-Moreno D, Catalán J, Cross RK, Stockmann-Juvala H, Cabellos J, Lopes VR, Matzke M, Ferraz N, Izquierdo JJ, Navas JM, Park M, Svendsen C, Janer G. Quality evaluation of human and environmental toxicity studies performed with nanomaterials – the GUIDEnano approach. *Environmental Science: Nano*, 2018, 5, 381-397. <https://doi.org/10.1039/C7EN00716G>
- Gopakumar AM, Balachandran PV, Xue D, et al. Multi-objective Optimization for Materials Discovery via Adaptive Design. *Sci Rep* 8, 3738 (2018). <https://doi.org/10.1038/s41598-018-21936-3>
- Gupta, SC. (2022). Machine Learning vs. Traditional Software Development. LinkedIn. <https://www.linkedin.com/pulse/machine-learning-vs-traditional-software-development-ml4devs-gupta/>
- Le Novère, N., et al., *Minimum information requested in the annotation of biochemical models (MIRIAM)*. *Nature biotechnology*, 2005. **23**(12): p. 1509.
- Lynch, I., Afantitis, A., Exner, T., & Papadiamantis, A. (2023). WorldFAIR Project (D4.1) Nanomaterials domain-specific FAIRification mapping (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7887341>
- Lynch, I., Dawson, K.A. Protein-nanoparticle interactions, *Nano Today*, 2008, 3, 40-47. [https://doi.org/10.1016/S1748-0132\(08\)70014-8](https://doi.org/10.1016/S1748-0132(08)70014-8)
- Lynch, I., Afantitis, A., Exner, T., & Papadiamantis, A. (2023). WorldFAIR Project (D4.1) Nanomaterials domain-specific FAIRification mapping (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7887341>

- Marchese Robinson RL, Lynch I, Peijnenburg W, Rumble J, Klaessig F, Marquardt C, Rauscher H, Puzyn T, Purian R, Åberg C, Karcher S, Vriens H, Hoet P, Hoover MD, Hendren CO, Harper SL. How Should the Completeness and Quality of Curated Nanomaterial Data Be Evaluated? *Nanoscale* 2016, 8 (19), 9919–9943. <https://doi.org/10.1039/C5NR08944A>.
- Rawashdeh, S. AI's mysterious 'black box' problem, explained. (2023), Short article: <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>
- Shao G, Beronius A, Nymark P. SciRAPnano: a pragmatic and harmonized approach for quality evaluation of in vitro toxicity data to support risk assessment of nanomaterials. *Front. Toxicol.*, 2023, 17. <https://doi.org/10.3389/ftox.2023.1319985>
- Sosnowska, A., Bulawska, N., Kowalska, D., & Puzyn, T. Towards higher scientific validity and regulatory acceptance of predictive models for PFAS. *Green Chemistry*, 2023, 25(4), 1261-1275. <https://doi.org/10.1039/D2GC04341F>
- Subbotina, J., & Lobaskin, V. Multiscale modeling of bio-nano interactions of zero-valent silver nanoparticles. *J. Phys. Chem. B*, 126(6), 1301-1314. (2022). <https://doi.org/10.1021/acs.jpcc.1c09525>
- van de Sandt S, Dallmeier-Tiessen S, Lavasa A, Petras V. The definition of reuse. *Data Sci J* (2019) 18:22. <https://doi.org/10.5334/dsj-2019-022>
- van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014, 14 (1), 1144. <https://doi.org/10.1186/1471-2458-14-1144>.
- Varsou DD, Afantitis A, Tsoumanis A, Melagraki G, Sarimveis H, Valsami-Jones E, Lynch I. A safe-by-design tool for functionalised nanomaterials through the Enalos Nanoinformatics Cloud platform. *Nanoscale Advances*, 2019, 1, 706-718. <https://doi.org/10.1039/C8NA00142A>
- Varsou DD, Kolokathis PD, Antoniou M, Sidiropoulos N, Tsoumanis A, Melagraki G, Lynch I, Afantitis A. In Silico Assessment of Nanoparticle Toxicity Powered by Enalos Cloud Platform: Integrating Machine Learning and Synthetic Data for Enhanced Nanosafety Evaluation. 2024, Submitted.
- Waltemath, D., et al., *Minimum information about a simulation experiment (MIASE)*. *PLoS computational biology*, 2011. 7(4): p. E1001122.

- Wheeler KE, Chetwynd AJ, Fahy KM, Hong BS, Tochihuitl JA, Foster LA, Lynch I. Environmental dimensions of the protein corona. *Nature Nanotechnology* 2021, 16, 617–629. <https://doi.org/10.1038/s41565-021-00924-1>
- Worth, A. P., & Balls, M. The principles of validation and the ECVAM validation process. *Alternatives to laboratory animals*, 30(2_suppl), 15-21. (2004). [10.1177/026119290403201s105](https://doi.org/10.1177/026119290403201s105)
- Wyrzykowska E, Mikolajczyk A, Lynch I, Jeliaskova N, Kochev N, Sarimveis H, Doganis P, Karatzas P, Afantitis A, Melagraki G, Serra A, Greco D, Subbotina J, Lobaskin V, Bañares MA, Valsami-Jones E, Jagiello K, Puzyn T. Representing and describing nanomaterials in predictive nanoinformatics. *Nat Nanotechnol.* 2022, 17(9):924-932. <https://doi.org/10.1038/s41565-022-01173-6>.

Appendix 1: NanoPharos FAIR Implementation Profile

This Appendix presents the FIP that has been produced for the NanoPharos database. The NanoPharos database has been developed by NovaMechanics Ltd. and offers users open access high-quality datasets in a ready-for-modelling tabular format. In this way, users are able to directly import these datasets into their workflows for processing and model development. NanoPharos is a new database, which is in the process of FAIRification based on the FAIR data principles and the FAIR data principle *interpretations* of the GoFAIR Foundation⁵⁵. Currently all datasets are assigned unique identifiers and linked to the respective publications that describe their metadata and are publicly available based on a clearly defined licence. There are plans to extend the services with the introduction of communication protocols to allow machine findability and data retrieval and submission. Furthermore, these will be coupled with the integration into NanoPharos of relevant ontologies to ensure that all data and metadata are machine understandable and actionable. In this way, it will be possible to link the data and metadata of the datasets with the respective models that use these datasets increasing the FAIRification level of the data-driven models. The current version of the NanoPharos FIP (V1.0) is presented below.

NanoPharos | NanoPharos Database: Ready-for-Modelling Datasets for Nanoinformatics

Organization	GO FAIR
Created by	Anastasios Papadiamantis
Based on	FIP Wizard 4, 4.0.3 (gofair:fip-wizard-4:4.0.3)
Project	Defining FAIR Implementation Profile
Phase	
Project Tags	Type: FIP
Description	NanoPharos Database: Ready-for-Modelling Datasets for Nanoinformatics
Created at	09 Oct 2023

I. About

No questions

II. Declare your FAIR Implementation Community

1. Select your FAIR Implementation Community

⁵⁵ <https://www.gofair.foundation/interpretation>



European Nanotechnology Data, Knowledge and Informatics Community

Started by the NanoCommons infrastructure project, the NanoCommons community infrastructure is meant to be the data and knowledge hub of the EU NanoSafety Cluster coordinating nanosafety research in Europe. Led by three core partners being part first in the NanoCommons project and now in the WorldFAIR project, it is connecting all stakeholders from current nanomaterials and new advance materials projects to foster knowledge exchange and harmonisation of the data ecosystem.

- [See more here](#)

(Nanopublication:

<http://purl.org/np/RAWg5p8lqzdX0PIYNBpze1iUpyhHxD7JgDQKWxsFNhSZk#NanoCommons>)

2. Who is the Community Data Steward?

✓ 0000-0002-1297-3104

3. Select the type of digital object you are focusing on in this FIP

! *This question has not been answered yet!*

4. If you have, please provide an accessible identifier of one of your digital objects of the chosen type.

! *This question has not been answered yet!*

5. Link to a case study

! *This question has not been answered yet!*

6. Specify the start date for the validity of the FIP

✓ 2023-10-06

7. Specify the end date for the validity of the FIP

✓ 2024-10-31

III. Declarations for Findability

1. Declaration F1 Metadata: What globally unique, persistent, resolvable identifier service do you use for metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

1.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

1.b.1.a.1. Select the FAIR Enabling Resource



DOI | Digital Object Identifier

The digital object identifier (DOI) system originated in a joint initiative of three trade associations in the publishing industry (International Publishers Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers). The system was announced at the Frankfurt Book Fair 1997. The International DOI Foundation (IDF) was created to develop and manage the DOI system, also in 1997. The DOI system was adopted as International Standard ISO 26324 in 2012. The DOI system implements the Handle System and adds a number of new features. The DOI system provides an infrastructure for persistent unique identification of objects of any type. The DOI system is designed to work over the Internet. A DOI name is permanently assigned to an object to provide a resolvable persistent network link to current information about that object, including where the object, or information about it, can be found on the Internet. While information about an object can change over time, its DOI name will not change. A DOI name can be resolved within the DOI system to values of one or more types of data relating to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers and descriptive metadata. The DOI system enables the construction of automated services and transactions. Applications of the DOI system include but are not limited to managing information and documentation location and access; managing metadata; facilitating electronic transactions; persistent unique identification of any form of any data; and commercial and non-commercial transactions. The content of an object associated with a DOI name is described unambiguously by DOI metadata, based on a structured extensible data model that enables the object to be associated with metadata of any desired degree of precision and granularity to support description and services. The data model supports interoperability between DOI applications. The scope of the DOI system is not defined by reference to the type of content (format, etc.) of the referent, but by reference to the functionalities it provides and the context of use. The DOI system provides, within

networks of DOI applications, for unique identification, persistence, resolution, metadata and semantic interoperability.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAnAWGdel_1GGmDAqv-vZjby5XqbL2ZujNz1vgwK_6cRI#DOI)

1.b.1.a.2. This implementation choice is:

- ✓ a. Currently in use by the community

1.b.1.a.3. Implementation Consideration (optional)

- ✓ The implementation was made on the data-related needs of the nanosafety and nanomaterials community for high-quality datasets in a ready-for-modelling format for direct import into computational workflows.

2. Declaration F1 Data: What globally unique, persistent, resolvable identifier service do you use for datasets?

- ✓ b. Declaration: FAIR Enabling Resource(s)

2.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

2.b.1.a.1. Select the FAIR Enabling Resource



DOI | Digital Object Identifier

The digital object identifier (DOI) system originated in a joint initiative of three trade associations in the publishing industry (International Publishers Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers). The system was announced at the Frankfurt Book Fair 1997. The International DOI Foundation (IDF) was created to develop and manage the DOI system, also in 1997. The DOI system was adopted as International Standard ISO 26324 in 2012. The DOI system implements the Handle System and adds a number of new features. The DOI system provides an infrastructure for persistent unique identification of objects of any type. The DOI system is designed to work over the Internet. A DOI name is permanently

assigned to an object to provide a resolvable persistent network link to current information about that object, including where the object, or information about it, can be found on the Internet. While information about an object can change over time, its DOI name will not change. A DOI name can be resolved within the DOI system to values of one or more types of data relating to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers and descriptive metadata. The DOI system enables the construction of automated services and transactions. Applications of the DOI system include but are not limited to managing information and documentation location and access; managing metadata; facilitating electronic transactions; persistent unique identification of any form of any data; and commercial and non-commercial transactions. The content of an object associated with a DOI name is described unambiguously by DOI metadata, based on a structured extensible data model that enables the object to be associated with metadata of any desired degree of precision and granularity to support description and services. The data model supports interoperability between DOI applications. The scope of the DOI system is not defined by reference to the type of content (format, etc.) of the referent, but by reference to the functionalities it provides and the context of use. The DOI system provides, within networks of DOI applications, for unique identification, persistence, resolution, metadata and semantic interoperability.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAnAWGdel_1GGmDAqv-vZjby5Xqbl2ZujNz1vgwK_6cRI#DOI)

2.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

2.b.1.a.3. Implementation Consideration (optional)

✓ Datasets are linked to their metadata via their respective peer-reviewed publications.

3. Declaration F2: What metadata schema do you use for findability?

✓ a. Declaration: No implementation choice has been made by this community

3.a.1. Considerations (optional)

✓ As of now, no specific implementation has been made due to the different requirements of the computational scientists / modellers.

4. Declaration F3: What is the schema that links the persistent identifiers of your data to the metadata description?

✓ b. Declaration: FAIR Enabling Resource(s)

4.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

4.b.1.a.1. Select the FAIR Enabling Resource

**DOI | Digital Object Identifier**

The digital object identifier (DOI) system originated in a joint initiative of three trade associations in the publishing industry (International Publishers Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers). The system was announced at the Frankfurt Book Fair 1997. The International DOI Foundation (IDF) was created to develop and manage the DOI system, also in 1997. The DOI system was adopted as International Standard ISO 26324 in 2012. The DOI system implements the Handle System and adds a number of new features. The DOI system provides an infrastructure for persistent unique identification of objects of any type. The DOI system is designed to work over the Internet. A DOI name is permanently assigned to an object to provide a resolvable persistent network link to current information about that object, including where the object, or information about it, can be found on the Internet. While information about an object can change over time, its DOI name will not change. A DOI name can be resolved within the DOI system to values of one or more types of data relating to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers and descriptive metadata. The DOI system enables the construction of automated services and transactions. Applications of the DOI system include but are not limited to managing information and documentation location and access; managing metadata; facilitating electronic transactions; persistent unique identification of any form of any data; and commercial and non-commercial transactions. The content of an object associated with a DOI name is described unambiguously by DOI metadata, based on a structured extensible data model that enables the object to be associated with metadata of any desired degree of precision and granularity to support description and services. The data model supports interoperability between DOI applications. The scope of the DOI system is not defined by reference to the type of content (format, etc.) of the referent, but by reference to the functionalities it provides and the context of use. The DOI system provides, within networks of DOI applications, for unique identification, persistence, resolution, metadata and semantic interoperability.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAnAWGdel_1GGmDAqv-vZjby5XqbL2ZujNz1vgwK_6cRI#DOI)

4.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

4.b.1.a.3. Implementation Consideration (optional)

✓ Datasets are linked to their metadata via their respective peer-reviewed publications.

5. Declaration F4 Metadata: Which service do you use to publish your metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

5.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

5.b.1.a.1. Select the FAIR Enabling Resource



NanoPharos Database: Ready-for-Modelling Datasets for Nanoinformatics

NanoPharos is a database providing high-quality datasets in a ready-for-modelling format for direct import in nanoinformatics (i.e., machine learning, AI) workflows.

- [See more here](#)

(Nanopublication:

<http://purl.org/np/RA-7YdatF38SxYdikEs11XiD9rEpYOsXZaTu8o9d1vRp0#NanoPharos>)

5.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

5.b.1.a.3. Implementation Consideration (optional)

✓ The NanoPharos database provides access to ready-for-modelling datasets for direct import into computational workflows, which are directly linked to their respective peer-reviewed publications for metadata retrieval.

6. Declaration F4 Datasets: Which service do you use to publish your datasets?

✓ b. Declaration: FAIR Enabling Resource(s)

6.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

6.b.1.a.1. Select the FAIR Enabling Resource



NanoPharos Database: Ready-for-Modelling Datasets for Nanoinformatics

NanoPharos is a database providing high-quality datasets in a ready-for-modelling format for direct import in nanoinformatics (i.e., machine learning, AI) workflows.

- [See more here](#)

(Nanopublication:

<http://purl.org/np/RA-7YdatF38SxYdikEs11XiD9rEpYOsXZaTu8o9d1vRp0#NanoPharos>)

6.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

6.b.1.a.3. Implementation Consideration (optional)

✓ The NanoPharos database provides public access to ready-for-modelling datasets for direct import into computational workflows, which are directly linked to their respective peer-reviewed publications for metadata retrieval.

IV. Declarations for Accessibility

1. Declaration A1.1 Metadata: Which standardized communication protocol do you use for metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

1.b.1. List the FAIR Enabling Resource(s)

Answers (2 items)

1.b.1.a.1. Select the FAIR Enabling Resource

**HTTPS | Hypertext Transfer Protocol Secure**

Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The protocol is therefore also referred to as HTTP over TLS, or HTTP over SSL

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAF1ANn-BCFop0OBMOC7S8NtG0y_xYhRX4tAu37XZVCo0#HTTPS)

1.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

1.b.1.a.3. Implementation Consideration (optional)

✓ The NanoPharos database is publicly accessible via it's dedicated URL.

1.b.1.b.1. Select the FAIR Enabling Resource

**REST | Representational state transfer**

REST defines a set of constraints for how the architecture of an Internet-scale distributed hypermedia system, such as the Web, should behave.

- [See more here](#)

(Nanopublication:

<http://purl.org/np/RAszH6lU-Zc3UO7MHPKj1Lb0dmMmaTJrRvQ0jqpXMyFY4#REST>)

1.b.1.b.2. This implementation choice is:

✓ c. Is planned to be used in the future

1.b.1.b.3. Implementation Consideration (optional)

✓ A REST API will be implemented for retrieving and submitting data to the database.

2. Declaration A1.1 Datasets: Which standardized communication protocol do you use for datasets?

✓ b. Declaration: FAIR Enabling Resource(s)

2.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

2.b.1.a.1. Select the FAIR Enabling Resource

✓

HTTPS | Hypertext Transfer Protocol Secure

Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The protocol is therefore also referred to as HTTP over TLS, or HTTP over SSL

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAF1ANn-BCFop0OBMOC7S8NtG0y_xYhRX4tAu37XZVCo0#HTTPS)

2.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

2.b.1.a.3. Implementation Consideration (optional)

✓ Datasets are publicly available for access and download via a dedicated link.

3. Declaration A1.2 Metadata: Which authentication & authorisation service do you use for metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

3.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

3.b.1.a.1. Select the FAIR Enabling Resource

✓

HTTPS | Hypertext Transfer Protocol Secure

Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The protocol is therefore also referred to as HTTP over TLS, or HTTP over SSL

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAF1ANn-BCFop0OBMOC7S8NtG0y_xYhRX4tAu37XZVCo0#HTTPS)

3.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

3.b.1.a.3. Implementation Consideration (optional)

✓ No specific authentication & authorisation services are currently used as the datasets are publicly available and linked to the respective publications. Direct links to the peer-reviewed publications are provided for metadata retrieval, but these are handled by the journals.

4. Declaration A1.2 Datasets: Which authentication & authorisation service do you use for datasets?

✓ b. Declaration: FAIR Enabling Resource(s)

4.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

4.b.1.a.1. Select the FAIR Enabling Resource



HTTPS | Hypertext Transfer Protocol Secure

Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The protocol is therefore also referred to as HTTP over TLS, or HTTP over SSL

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAF1ANn-BCFop0OBMOC7S8NtG0y_xYhRX4tAu37XZVCo0#HTTPS)

4.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

4.b.1.a.3. Implementation Consideration (optional)

✓ No specific authentication & authorisation services are currently used as the datasets are publicly available and linked to the respective publications. Direct links to the peer-reviewed publications are provided for metadata retrieval, but these are handled by the journals.

5. Declaration A2: What metadata preservation policy do you use?

✓ a. Declaration: No implementation choice has been made by this community

5.a.1. Considerations (optional)

! *This question has not been answered yet!*

V. Declarations for Interoperability

1. Declaration I1 Metadata: What knowledge representation language (allowing machine interoperation) do you use for metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

1.b.1. List the FAIR Enabling Resource(s)

Answers (4 items)

1.b.1.a.1. Select the FAIR Enabling Resource

✓

****CSV File Format ****

Files with .csv (Comma Separated Values) extension represent plain text files that contain records of data with comma separated values. Each line in a CSV file is a new record from the set of records contained in the file. Such files are generated when data transfer is intended from one storage system to another. Since all applications can recognize records separated by comma, import of such data files to database is done very conveniently. Almost all spreadsheet applications such as Microsoft Excel or OpenOffice Calc can import CSV without much effort. Data imported from such files is arranged in cells of a spreadsheet for representation to user.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAcNC9zQOEc9RPXD07R91aMNSlYmJR9G4kRT_Pm4FpzKE#CSV)

1.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

1.b.1.a.3. Implementation Consideration (optional)

! *This question has not been answered yet!*

1.b.1.b.1. Select the FAIR Enabling Resource



OWL | Web Ontology Language

The Web Ontology Language (OWL) is a family of knowledge representation languages or ontology languages for authoring ontologies or knowledge bases. The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium (W3C) and has attracted academic, medical and commercial interest. The OWL 2 Web Ontology Language, informally OWL 2, is an ontology language for the Semantic Web with formally defined meaning. OWL 2 ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL 2 ontologies can be used along with information written in RDF, and OWL 2 ontologies themselves are primarily exchanged as RDF documents.

- [See more here](#)

(Nanopublication: http://purl.org/np/RAIpnldLlledp5J7Jcy8pt_X9_YpOez4rO-fHxZI0T96Y#OWL)

1.b.1.b.2. This implementation choice is:



c. Is planned to be used in the future

1.b.1.b.3. Implementation Consideration (optional)

! *This question has not been answered yet!*

1.b.1.c.1. Select the FAIR Enabling Resource



RDF | Resource Description Framework

The Resource Description Framework (RDF) is a framework for representing information in the Web.

- [See more here](#)

(Nanopublication:

<http://purl.org/np/RAutRQwoS4d5eLq7eBV1xsnWZ2spDYH4xfhhRzOxSZdhs#RDF>)

1.b.1.c.2. This implementation choice is:

✓ c. Is planned to be used in the future

1.b.1.c.3. Implementation Consideration (optional)

! *This question has not been answered yet!*

1.b.1.d.1. Select the FAIR Enabling Resource



JSON-LD | JavaScript Object Notation for Linking Data

JSON-LD is a JSON-based format to serialize Linked Data. The syntax is designed to easily integrate into deployed systems that already use JSON, and provides a smooth upgrade path from JSON to JSON-LD. It is primarily intended to be a way to use Linked Data in Web-based programming environments, to build interoperable Web services, and to store Linked Data in JSON-based storage engines. JSON-LD is a concrete RDF syntax. A JSON-LD document is both an RDF document and a JSON document and correspondingly represents an instance of an RDF data model. However, JSON-LD also extends the RDF data model to optionally allow JSON-LD to serialize generalized RDF Datasets.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAQKjgd7Ug9xSo4J0REW_AHGOJyaF9-ydj60nungQ0qVg#JSON-LD)

1.b.1.d.2. This implementation choice is:

✓ c. Is planned to be used in the future

1.b.1.d.3. Implementation Consideration (optional)

! *This question has not been answered yet!*

2. Declaration I1 Datasets: What knowledge representation language (allowing machine interoperation) do you use for datasets?

✓ a. Declaration: Chemical structure representation such as InChI

2.a.1. Considerations (optional)

! *This question has not been answered yet!*

3. Declaration I2 Metadata: What structured vocabulary do you use to annotate your metadata records?

✓ a. Declaration: No implementation choice has been made by this community

3.a.1. Considerations (optional)

! *This question has not been answered yet!*

4. Declaration I2 Datasets: What structured vocabulary do you use to encode your datasets

✓ a. Declaration: No implementation choice has been made by this community

4.a.1. Considerations (optional)

! *This question has not been answered yet!*

5. Declaration I3 Metadata: What semantic model do you use for your metadata records?

✓ a. Declaration: No implementation choice has been made by this community

5.a.1. Considerations (optional)

! *This question has not been answered yet!*

6. Declaration I3 Datasets: What semantic model do you use for your datasets?

✓ a. Declaration: No implementation choice has been made by this community

6.a.1. Considerations (optional)

! *This question has not been answered yet!*

VI. Declarations for Reusability

1. Declaration R1.1 Metadata: Which usage license do you use for your metadata records?

✓ b. Declaration: FAIR Enabling Resource(s)

1.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

1.b.1.a.1. Select the FAIR Enabling Resource



CC BY 4.0 | Attribution 4.0 International

Using this licence you are free to share and adapt the resource but you must give appropriate credit.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAQ_sGdY_Qc7l1O_zmn4nr-pMBOxKU04Ur9s998rS6Fc#CC-BY-4.0)

1.b.1.a.2. This implementation choice is:

✓ a. Currently in use by the community

1.b.1.a.3. Implementation Consideration (optional)

✓ Metadata are linked to their respective peer-reviewed publications, thus the licensing is handled by these. Any metadata available in NanoPharos are available by a CC-BY 4.0 license.

2. Declaration R1.1 Datasets: Which usage license do you use for your datasets?

✓ b. Declaration: FAIR Enabling Resource(s)

2.b.1. List the FAIR Enabling Resource(s)

Answers (1 items)

2.b.1.a.1. Select the FAIR Enabling Resource



CC BY 4.0 | Attribution 4.0 International

Using this licence you are free to share and adapt the resource but you must give appropriate credit.

- [See more here](#)

(Nanopublication:

http://purl.org/np/RAQ_sGdY_Qc7l1O_zmn4nr-pMBOxKU04Ur9s998rS6Fc#CC-BY-4.0)

2.b.1.a.2. This implementation choice is:



a. Currently in use by the community

2.b.1.a.3. Implementation Consideration (optional)



This question has not been answered yet!

3. Declaration R1.2 Metadata: What metadata schema do you use for describing the provenance of your metadata records?



a. Declaration: No implementation choice has been made by this community

3.a.1. Considerations (optional)



This question has not been answered yet!

4. Declaration R1.2 Datasets: What metadata schema do you use for describing the provenance of your datasets?



a. Declaration: No implementation choice has been made by this community

4.a.1. Considerations (optional)



This question has not been answered yet!

5. Declaration R1.3: Your community uses this FAIR Implementation Profile to link to domain-relevant community standards. Please acknowledge this statement by clicking on 'Read and understood'.

✓ a. Read and understood.

VII. Register a new resource as a nanopublication

No questions