LIVERPOOL JOHN MOORES UNIVERSITY

LJMU Research Online

**Cronin, MTD, Basiri, H, Belfield, SJ, Chavan, S, Chrysochoou, G, Enoch, SJ, Firman, JW, Gomatam, A, Hardy, B, Helmke, PS, Madden, JC, Maran, U, March-Vila, E, Nikolov, NG, Pastor, M, Piir, G, Popelier, PLA, Sild, S, Smajić, A, Sp̂înu, N and Wedebye, EB**

# The Findable, Accessible, Interoperable, Reusable (FAIR) Lite Principles to ensure utility of computational toxicology models

**https://researchonline.ljmu.ac.uk/id/eprint/27381/**

**Article**

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Cronin, MTD ORCID logoORCID: https://orcid.org/0000-0002-6207-4158, Basiri, H, Belfield, SJ, Chavan, S, Chrysochoou, G, Enoch, SJ ORCID logoORCID: https://orcid.org/0000-0001-9111-5783, Firman, JW ORCID logoORCID: https://orcid.org/0000-0003-0319-1407, Gomatam, A, Hardy, B,**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

*ALTEX, accepted manuscript*
*published September 25, 2025*
*doi:10.14573/altex.2502021*

Research Article

# The Findable, Accessible, Interoperable, Reusable (FAIR) Lite Principles to Ensure Utility of Computational Toxicology Models

*Mark T. D. Cronin[1], Homa Basiri[1], Samuel J. Belfield[1,2], Swapnil Chavan[3], Georgios Chrysochoou[1], Steven J. Enoch[1], James W. Firman[1], Anish Gomatam[1], Barry Hardy[4], Palle S. Helmke[5], Judith C. Madden[1], Uko Maran[6], Eric March-Vila[7], Nikolai G. Nikolov[8], Manuel Pastor[7], Geven Piir[6], Paul L. A. Popelier[2], Sulev Sild[6], Aljoša Smajić[5], Nicoleta Spînu[1] and Eva B. Wedebye[8]*

[1]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK; [2]Department of Chemistry, University of Manchester, Manchester, UK; [3]Unit of Chemical and Pharmaceutical Toxicology, Research Institutes of Sweden (RISE), Södertalje, Sweden; [4]Edelweiss Connect, Technology Park Basel, Basel, Switzerland; [5]Department of Pharmaceutical Sciences, University of Vienna, Vienna, Austria; [6]Institute of Chemistry, University of Tartu, Tartu, Estonia; [7]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra, Barcelona, Spain; [8]Technical University of Denmark, National Food Institute, Research Group for Chemical Risk Assessment and GMO, Kgs. Lyngby, Denmark

**Abstract**

A broad range of computational models are available for animal-free chemical safety assessment. The models are used to predict a variety of endpoints, including adverse effects or apical endpoints, toxicokinetic properties and exposure, often from chemical structure or *in vitro* inputs alone. To support their wider use, such models need to be Findable, Accessible, Interoperable, Reusable (FAIR). This study has reevaluated the existing FAIR principles applied to quantitative structure-activity relationships (QSARs) in order to adapt these principles to a wider range of computational models. Despite the breadth and variety of approaches, many computational models comprise common components including the training series, information about the modelling engine and the model itself. As a result, a refined set of four FAIR Lite principles is proposed based on the methodological foundations of computational toxicology which are unambiguously understood by practitioners such as developers and end-users. To this end, it is proposed that to comply with the original , a computational toxicology model should be associated with (i) a globally unique identifier for model citation; (ii) the capture and curation of the model; (iii) the metadata for the dependent and independent variables and, where possible, data; and (iv) storage in a searchable and interoperable platform. The FAIR Lite principles are mapped onto the original FAIR principles applied to QSARs, thereby demonstrating that a simpler checklist approach covers all aspects.

**Plain language summary**

Many types of computational models are used in animal-free chemical safety assessment. These are utilized to make predictions for numerous endpoints, primarily focusing on the hazardous properties of, or exposure to, a chemical substance. The models use information from chemical structures and/ or properties, or other non-animal data as inputs. It is essential that the risk assessor or toxicologist can find and utilize the models with confidence. The previously developed Findable, Accessible, Interoperable and Reusable (FAIR) principles for computational models are a framework intended to ensure that models are accessible and stored appropriately. This investigation has refined the original FAIR principles applied to computational models to capture information for all types of modelling approaches that may be used in chemical safety assessment. The new principles, termed FAIR Lite, encapsulate the original principles in four criteria relating to identifiers, description of a model, its (meta)data and storage.

## Introduction

The move to modernize chemical safety assessment, without the overt reliance on animals, will depend on a number of technologies. Amongst the new approaches, computational models are finding use as part of the replacement strategy of the

Received February 2, 2025; Accepted September 23, 2025;
Epub September 25, 2025; © The Authors, 2025.

**ALTEX 42(4), ###-###.** doi:10.14573/altex.2502021

Correspondence: Mark Cronin, PhD
School of Pharmacy and Biomolecular Sciences,
Liverpool John Moores University,
Byrom Street, Liverpool L3 3AF, UK
(m.t.cronin@ljmu.ac.uk)

traditional animal tests for hazard identification (either directly or as part of a weight-of-evidence), estimation of exposure to chemicals and substances, as well as the utilization of *in vitro*, *in chemico* and -omics data (Barber et al., 2024). Computational approaches will also be fundamental building blocks of tiered testing strategies including, but not limited to, Integrated Approaches to Testing and Assessment (IATA) (Delrue et al., 2016), Defined Approaches (DAs) (Macmillan and Chilton, 2019) and Next Generation Risk Assessment (NGRA) (Baltazar et al., 2020) as well as approaches such as the ASPIS Safety Profiling Algorithm (ASPA) (Luijten, 2024), which attempts to provide a workflow as the backbone to implementing NGRA.

There is an exceptionally broad range of types of computational models that may be applied in NGRA, and hence in tiered strategies such as ASPA (Madden et al., 2020; Westmoreland et al., 2022). Table 1 summarizes a variety of computational models that have been used to interpret, extrapolate and predict the broad spectrum of data required for chemical safety assessment. Typically, these models have been derived from existing experimental data, usually taken from historical testing or measurements, and are used to make estimates for data gaps. Such models, which may be termed *in silico*, or computational, make predictions related to the toxicodynamic interactions and toxicokinetic effects (or both) of chemicals relating, in part at least, to the hazard and internal exposure of the chemical (illustrative endpoints are also summarized in Table 1). To facilitate and support model development, there is also a need to have access to data sources. There are currently several initiatives for toxicological data and information sharing, such as the United States Environmental Protection Agency (US EPA) CompTox Chemicals Dashboard[1] (Williams et al., 2017) which incorporates diverse data such as ToxValDB and invitroDB; COSMOS Next Generation[2] (Yang et al., 2021); and Open TG-Gates (Igarashi et al., 2015). Whilst there are great advantages in sharing data, there are also many technological and business-related (i.e. confidentiality) challenges (Pastor et al., 2018). The issues with aggregating human health risk assessment-relevant information on chemicals, with a particular focus on available datasets and databases were investigated by Freudenthal et al. (2024). The outcome of their investigation emphasized the requirement for all stakeholders to improve aspects of data sharing related to data harmonization and transparency, amongst other aspects.

To allow and encourage the use of computational models in non-animal chemical safety assessment, there is a need to facilitate access to the models. This is encapsulated in the Findable, Accessible, Interoperable, Reusable (FAIR) principles for data sharing (Wilkinson et al., 2016). The FAIR principles are also applicable to supporting the sustainability and reusability of research software (FAIR4RS) (Barker et al., 2022). FAIR4RS demonstrated that FAIR was applicable to digital objects beyond data themselves and provided the basis for adaptation of the FAIR principles for data management to software. The implementation of FAIR4RS enables and encourages sharing of software, as well as maximizing its impact. The FAIR and FAIR4RS principles were adapted for *in silico* models in toxicology, notably for QSARs by Cronin et al. (2023) with a full evaluation by Belfield et al. (2025). Whilst the FAIR principles have found application for QSARs (Belfield et al., 2025), the types of modelling approaches and data utilized in all aspects of computational toxicology are much broader, as illustrated in Table 1. Despite enormous diversity in models and modelling approaches, all data-driven and/or knowledge-based computational toxicology models share similar characteristics (there will be exceptions for simulations). Cronin et al. (2023) defined these characteristics in terms of the training series, a modelling engine and the model itself. The training workflow uses the modelling engine to provide the model. In contrast, the prediction workflow inputs a chemical into a modelling engine to make a prediction that provides new knowledge about the chemical. To consider the full suite of computational toxicology models in Table 1, the training series, model and modelling engine can be rationalized into four components:

- Training series: dependent variables (those that are modelled);
- Training series: independent variables (those that affect or influence the dependent variables);
- Modelling engine: the statistical or other algorithm that allows for the creation of the predictive model;
- Model: the parameters of the modelling engine resulting from it being exposed to the training series (this may be thought of as the algorithm or mathematical formula that is applied).

At the current time, there is no single repository, or inventory, of computational toxicology models. At best, models are scattered throughout the historical scientific literature and some of the resources described by Belfield et al., (2025). The lack of a single location, or search engine, that would enable a researcher or risk assessor to find a suitable model to predict a particular endpoint or property, for a specific substance or group of substances is restrictive to the implementation of models. These issues, as well as aspects of data integration and sharing are fundamental to the replacement of animals for chemical safety assessment (Westmoreland et al., 2022; Cronin et al., 2025). Thus, to support non-animal chemical safety assessment, there is a clear requirement that all relevant computational toxicological models must be retrievable and usable, in other words they should be FAIR. This need goes beyond applications solely to QSARs to the broad range of models, some of which are summarized in Table 1. The purpose of this paper was to further enhance the existing FAIR principles (Wilkinson et al., 2016), their adaptation to research software (Barker et al., 2022) and, specifically for computational toxicology, QSAR models (Belfield et al., 2025) to provide a practical solution to supporting FAIR models. The consideration here is that a model is intrinsically different to, and often more complex than, the data to be modelled, or the software used for model development or implementation. Given that computational toxicology models comprise similar core elements (i.e., the data, modelling engine and model), this paper aimed to evaluate a broad range of models with a pragmatic and helpful view to develop a practical, condensed checklist (termed FAIR Lite), with associated solutions, for making the models FAIR. The intention here is to adapt the FAIR principles, applied to QSARs, to broader methodological foundations of computational toxicology. In addition, the FAIR Lite principles intend to be a pragmatic, lightweight, flexible and adaptable solution to make computational toxicology models sustainable and reusable, going beyond the application of FAIR principles to data

[1] https://comptox.epa.gov/dashboard/
[2] https://ng.cosmosdb.eu/

**Tab. 1: A brief overview and non-exhaustive summary of the computational toxicology approaches that may be applied in non-animal chemical safety assessment**

| Computational Toxicology Modelling Approach (acronyms defined in the legend) | Brief Description and Characteristics | Endpoints Predicted | Indicative References and/or Reviews |
|---|---|---|---|
| Local QSAR | A statistical relationship between activity and molecular properties relating to a restricted chemical space and / or a single mechanism or mode of action (where definable) | Toxicological effects or regulatory endpoints; toxicokinetic and physico-chemical properties | Cherkasov et al., 2014 |
| Global QSAR | A statistical relationship between activity and molecular properties related to a broad chemical space and multiple (or unknown) mechanisms or modes of action | Toxicological effects or regulatory endpoints; toxicokinetic and physico-chemical properties | Madden et al., 2020 |
| Structural alerts, e.g. within a knowledge-based expert system or computational profiler | Formalized structural pattern providing knowledge of chemistry underpinning a cause for certain adverse events | Possibility of a molecule to elicit a toxicological effect | Yang et al., 2020 |
| qAOP | Development of quantitative relationships between Molecular Initiating Events (MIEs) and Key Events (KEs), or KEs and KEs, or KEs and adverse outcomes | Prediction of downstream KEs or adverse outcomes | Spînu et al., 2020 |
| QST | Integration of various models, based around systems biology, to estimate toxicodynamic interactions and, when combined with modelling of toxicokinetic effects, define a response at the molecular level to the exposure of a particular dose of a substance | Prediction of adverse outcomes, potentially following a specified exposure | Sturla et al., 2014 |
| IVIVE | Models allowing for the extrapolation of nominal concentrations inducing an effect in an *in vitro* experiment to exposure that causes an adverse effect in an *in vivo* system | *In vivo* effect at a particular dose extrapolated from an effect at a known *in vitro* dose | Wambaugh et al., 2018 |
| Reverse dosimetry | Models to estimate an external exposure based on measured internal concentrations | The external concentration that would be required to achieve an internal (*in vivo*) concentration | Wetmore et al., 2015 |
| PBK model | Multicompartment models that predict the concentration of a chemical after exposure in a particular internal organ or tissue | Concentration-time profile of a substance in an organ, following a specific exposure | Rowland et al., 2011 |
| Aggregate exposure | Models to calculate an individual total exposure to a substance from all sources and exposure routes | Maximal exposure to a substance | Safford et al., 2015 |
| Molecular Mechanics | Uses classical mechanics and forcefields to model the structure, energy, and dynamics of molecules based on atoms and bonds | Conformational stability, steric interactions and energy-minimized molecular structures relevant to toxicodynamics | Wang et al., 2019 |
| Molecular docking | Models that capture interactions between proteins and ligands at atomic level | Binding of ligands to biological macromolecules | Trisciuzzi et al., 2018 |
| Quantum Mechanics | Quantum chemical methods that model the electronic structure and properties of molecules | Reactivity, electronic properties, mechanisms of toxicological action (e.g., covalent binding) | Kostal, 2023 |
| MD simulations | Models that predict dynamic behavior of biomolecules over time | Time dependent properties i.e., changes in conformation, protein folding, protein-ligand binding affinity, etc. | De Vivo, 2016 |
| Knowledge graphs | Integration of toxicological data from multiple sources | Toxicological effects or regulatory endpoints; toxicokinetic and physico-chemical properties | Sepehri et al., 2025 |
| Active learning | Model trained on active selection of most informative data points | Toxicological effects or regulatory endpoints; toxicokinetic and physico-chemical properties | Nahal et al., 2024 |

QSAR: Quantitative Structure-Activity Relationship; qAOP: quantitative Adverse Outcome Pathway; QST: Quantitative Systems Toxicology; IVIVE: *In Vitro- In Vivo* Extrapolation; PBK: Physiologically-Based Kinetic; MD: Molecular Dynamics simulations

(Wilkinson et al., 2016) and research software (Barker et al., 2022). Thus, the FAIR Lite principles are intended to be easy to use, portable, transferable and flexible. In particular, they should be applicable to all computational toxicology models bearing in mind future technologies that may be developed, for example being derived from recent advances in AI and Large Language Models, as well as many of the probabilistic (e.g. Bayesian) decision-making and other machine learning (ML) models being applied.

## 2 Methods

### 2.1 Evaluation of Existing FAIR Principles Applied to QSARs

The 18 core FAIR principles applied to QSARs (which gave a total of 20 principles following sub-division) (Cronin et al., 2023; Belfield et al., 2025) were re-evaluated in this study. Using expert judgement, these principles were considered in the context of a broader range of computational toxicology models, as represented by those listed in Table 1. Specifically, the 13 model types identified in Table 1 were mapped against the four components identified above (dependent data, independent data, modelling engine and model) to identify their key features as relevant to the overall philosophy of the FAIR Principles. Commonalities in the models were identified, and the FAIR principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) were adapted and streamlined. The aim was to develop revised principles that would better meet the needs of computational toxicology models in non-animal chemical safety assessment.

### 2.2 Development of FAIR Lite Principles

Following evaluation of the FAIR principles applied to QSARs, in the context of the broader use of computational models in toxicology and for the purposes of risk assessment, fundamental and overriding concepts and principles were identified by expert judgement and consensus amongst the authors. These "essential" FAIR principles which can be applied to computational models are herein termed FAIR Lite. The FAIR Lite principles were created to cover the requirements to make models sharable and reusable. A checklist was developed to allow a model developer to assess whether they had met the FAIR Lite principles for *in silico* toxicology models.

### 2.3 Evaluation of FAIR Lite Principles

The FAIR Lite principles were evaluated in terms of the coverage compared to the broader, more detailed set of principles described by Cronin et al. (2023) and Belfield et al. (2025). The evaluation was performed manually by mapping each of the FAIR principles applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) onto the relevant FAIR Lite principles. The results are demonstrated in tabular form and graphically.

## 3 Results and Discussion

This study has investigated the possibility of ensuring that computational models for use in all aspects of (non-animal) chemical safety assessment are FAIR. The purpose was not to determine whether individual models were themselves FAIR, but rather if a common set of FAIR principles could be established relevant to the broad variety of computational toxicology models available. The various computational toxicology models can be defined according to their features relating to dependent and independent data, modelling engine and model – as described in Table 2. This exercise does not imply that all computational toxicology models are equivalent – this is clearly not the case – but rather that they comprise verifiable components. Since many computational models considered can be defined in terms of similar components, it would be logical to assume that some overarching considerations of the FAIR principles could be applied.

The motivation for making computational models for every aspect of chemical safety assessment FAIR is to provide researchers and risk assessors full access to the best and most appropriate model for a particular task. It is also motivated by the current difficulties in finding and utilizing (published) models, as well as the clear responsibility to ensure that the outputs (especially funded from public resources) are available for future use, for instance as report by Piir et al. (2018) with regard to the use of existing QSAR models for physicochemical properties, environmental fate, ecotoxicity, human health and toxicokinetics. It should also be noted that the FAIR principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025), or FAIR Lite principles are not intended to disqualify models with confidential or business-sensitive data, or commercial models where the algorithm is not disclosed; rather it is to allow data and models to be found and to be machine readable. The founding publication by Wilkinson et al. (2016) states that there are degrees of FAIRness and the modularity of the principles support a wide range of special circumstances, including highly sensitive or personally-identifiable data, and even non-data research objects. The purpose of these FAIR Lite principles is to follow good data management practices that allow the models to be found and used easily in a transparent and reproducible manner that may allow for acceptance. Indeed, for commercial models, ensuring and demonstrating models and other products are FAIR must be seen as a clear business opportunity. This is in concordance with interpretation and implementation of the FAIR principles with regard to "provenance", as described by Jacobsen et al. (Section 3.4.3; 2020). According to Jacobsen et al. (2020) provenance is interpreted as including how and why the resource was created, ownership etc. Further, the richness of the provenance, which we interpret as relating to the description and availability of "model", is related to actual reuse. It would be anticipated that commercial models for computational toxicology can be considered FAIR if they can supply relevant metadata to ensure machine readability. A minimal set of metadata should be agreed as an industry standard that could, for example, include the unique identifier, endpoint, property or effect modelled, the required input and intended output variables, applicability domain, storage and/or accessibility, without releasing the model's algorithm.

**Tab. 2: Mapping computational toxicology models to define the key features relevant to FAIR principles as defined in Cronin et al., 2023**

| Model type | Examples of dependent data | Examples of independent data | Modelling engine | Model |
|---|---|---|---|---|
| Local QSAR | Toxicity or adverse effects | Limited numbers of physico-chemical properties and/ or molecular descriptors | Statistical techniques suited to small data sets e.g. regression analysis, discriminant analysis | Regression or discriminant analysis equation |
| Global QSAR | Toxicity or adverse effects | Physico-chemical properties and / or molecular descriptors, fingerprints etc. | A broad range of ML or AI algorithms | Various, e.g. regression, classification, neural networks etc |
| Structural alerts, e.g. within a knowledge-based expert system or computational profiler | Toxicity or adverse effects, molecular initiating events, modes / mechanisms of toxic action | Sub-structural and/or functional group information in the form of structural fingerprints, ToxPrint chemotypes, etc. | A knowledge-based structural alert is derived from the expert-identified chemistry related to a toxicological event. Computationally derived alerts may be derived from the statistical analysis of structural fragments and their association with toxicity etc. | Identification of a structural alert in a molecule by comparison to a series of alerts in a ”profiler” or other software |
| qAOP | Relationship between MIEs, KEs and an adverse outcome | MIEs, KEs | Statistical or mechanistic modelling techniques | Quantitative model establishing a relationship between KEs and adverse outcomes |
| QST | Toxicodynamic interactions and toxicokinetic effects | Information on systems biology and specified exposure to a chemical | Multi-level models based on various statistical and empirical modelling techniques | Modelling of the effect of an exposure to a xenobiotic |
| IVIVE | *In vivo* biological response | Data from *in vitro* assays | Models for extrapolation based on pharmacokinetic principles | IVIVE model |
| Reverse dosimetry | Internal exposure measurements | External concentrations | Models for extrapolation of internal to external exposure | Quantitative modelling of internal to external – exposure relationship |
| PBK | Concentration of chemical in specific organ/tissue | Physiological and anatomical parameters; chemical-specific properties | Various types of multicompartmental models intending to simulate the passage, accumulation, metabolism etc. of a molecule within an organism | Comprehensive PBK model for chemical exposure assessment |
| Aggregate Exposure | Exposure levels estimation | Environmental concentrations | Statistical models aiming to combine multiple sources of exposure to a chemical allowing for an overall exposure to be calculated | Predictive model for the total exposure to a substance |
| Molecular Mechanics | Conformational changes, interaction energies | Atom types, bond lengths/angles, force field parameters | Classical force fields | Energy-minimized molecular structures |
| Molecular docking | Binding energy | 3D structures of ligands and target proteins | Scoring functions | Predicted protein-ligand binding pose and binding affinity |
| Quantum Mechanics | Electron density/charge distribution, reaction energies | Molecular geometries, basis sets, Hamiltonian operators | Probabilistic modelling based on Schrodinger’s equation and its approximations | Well-characterized electronic structure |
| MD simulations | Conformational changes, interaction energies | Initial 3D coordinates of biomolecules, temperature, solvent parameters | Physics-based simulation software | Simulation trajectory describing molecular motions and interactions |
| Knowledge graphs | Relationships between chemicals, biomolecules such as target proteins, associated pathways and outcomes | Data from different sources (e.g.: chemical and toxicological data, bioassay data, ontologies etc.) | Web technologies that integrate data using nodes and edges | A network representing toxicological knowledge gathered from multiple sources |
| Active learning | Toxicity or adverse effects | Physico-chemical properties and / or molecular descriptors, fingerprints etc. | Statistical techniques suited to the data | Regression or classification model trained on the most informative data points |

This investigation is based on internationally recognized FAIR principles, for instance those for data sharing (Wilkinson et al., 2016) and research software (Barker et al., 2022), but with specific consideration of their adaptation for toxicological QSARs (Cronin et al., 2023). A previous evaluation of the FAIR principles for toxicological QSARs (Belfield et al., 2025) found a number of areas where QSARs were lacking with regard to their being FAIR. These were the absence of a unique identifier and standardized ontologies for data and model description, not providing the data and allowing full interoperability with other software. These are likely to be similar to the broader range of models for computational toxicology and should be considered in depth.

### 3.1 "Fair Lite" Principles for Computational Toxicology Models

The FAIR principles applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) comprise 20 criteria as listed in Table 3. They were developed specifically with QSARs in mind. For the consideration of a wider variety of computational toxicology models, there is an opportunity to refine the principles to be all-encompassing. The evaluation of a broad range of computational toxicology models in Table 2 demonstrates that they can be evaluated according to the principles set out in Cronin et al. (2023), namely that they are based on data, involve a modelling approach (the so-called model engine) and that there is a resultant model. This provides an opportunity to broaden and simplify the FAIR principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) for any computational toxicology model. To this end, it is proposed that to comply with the original FAIR principles, **a computational toxicology model should demonstrate, or be associated with:**

1. **A globally unique identifier for model citation.** This is intended to assist in the findability of models and ensure that their description and data are identifiable.
2. **The capture and curation of the model.** This is essential for the interoperability and reusability of the model. Without an appropriate description the model cannot be used appropriately.
3. **The metadata for the dependent and independent variables and, if the chosen (commercial) license allows, data.** Description and, where appropriate, availability of data (for the training, test and validation sets) is vital for the transparency of the model, its findability and reusability. Data availability is subject to the sensitivity and confidentiality of the data.
4. **Storage of the model in a searchable and interoperable platform.** The ability to store, search for, and utilise, a particular model is essential for its accessibility and reusability.

These four principles are termed the "FAIR Lite" principles for computational toxicology models. The intention is to cover all relevant aspects of the original FAIR principles (Wilkinson et al., 2016) and FAIR4RS principles (Barker et al., 2022). It is recognized that models for computational toxicology offer unique challenges which, in some cases, extend these principles, whilst other principles may be less relevant or redundant. To demonstrate that all appropriate FAIR principles and criteria have been met, the original / adapted principles for QSARs have been mapped against the Wilkinson et al. (2016) FAIR principles (summarized in Tab. 3A and Fig. 1) and Barker et al. (2022) FAIR4RS principles (Tab. 3B and Fig. 2). It is clear that the four "FAIR Lite" principles capture all the criteria, with no missing principles. For mapping on FAIR4RS (Table 3B), some pragmatism was applied as the FAIR4RS principles were not intended to specifically capture models (hence no mention of models or algorithms) or the underlying data. In this regard, the FAIR Lite principles extend the coverage to include features that are essential for computational toxicology yet not assessed within FAIR4RS.

The FAIR Lite criteria provide a straightforward and fundamental means of evaluating the compliance, or otherwise, of a computational toxicology model with the FAIR principles. The first of the four criteria is the need for a globally unique identifier. Currently, few models are associated with such an identifier. Belfield et al. (2025) recommended using the Document Object Identifier (DOI) as it is well established for publications and retrievable on internet search engines. For example, the QsarDB repository provides DOIs for QSAR-related data publication and to individual QSAR models (Ruusmann et al., 2015). The second FAIR Lite criterion is the capture and curation of the model itself, preferably in a machine-readable format. This implies the description of the model which can include: (i) the training series and modelling engine, which allow for the model to be reproduced, and (ii) the set of parameters within the model (or algorithm / formula) generated by the training. An example of such digitalization and organization of data, and model representation in machine readable manner, is the QSAR DataBank archive format (Ruusmann et al., 2014), with other efforts reviewed in Sild et al. (2020). The capture and curation of a model will be also dependent on licensing and commercial considerations. Even if the model cannot be made available, it is in the interest of a vendor or supplier that the model is adequately documented and can be found and utilized. It is noted that representations of chemical structure such as SMILES and InChI may not be considered as persistent identifiers. Neither do they contain the 3D structural information that is available in, for instance, a .mol or .sdf file. The third criterion is that the metadata which describe the model are provided and, where possible, also the data on which the model is derived. The metadata must also include licensing information, which is necessary for users of the model to comply with restrictions on commercial use and licensing terms. (Meta)data need to be described in a consistent manner so that they are unambiguous, e.g., using a standardized ontology for toxicological tests and descriptor data (a summary of potential ontologies is provided in Table S1). In the context of computational toxicology, it is also important to distinguish between the licensing of model data and its (meta)data. Due to commercial interests and legal implications, restricted access to computational models might be reasonable. Therefore, an open license is highly recommended for the (meta)data to make the information about the models accessible and usable. Finally, according to the fourth criterion, models require a permanent, searchable storage solution where they can be used, downloaded, or linked to a model provider.

To make the FAIR Lite criteria into a practical scheme, Table 4 provides a checklist that could be used to evaluate whether any computational toxicology model complies with the principles and possible solutions to allow for compliance. It must be emphasized that neither the FAIR nor the FAIR Lite Principles are intended to validate a model, e.g. such as may be required for regulatory use, or determine data quality. For example, model validation, e.g., for regulatory acceptance of a prediction, should be performed separately, with approaches such as the implementation of the OECD Principles for the Validation of

**Tab. 3A: Mapping of the FAIR Principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) against the FAIR Lite Principles**

| FAIR Principles for *In Silico* Models and QSARs for toxicology (from Cronin et al., 2023; Belfield et al., 2025) | FAIR Lite 1: Unique Identifier | FAIR Lite 2: Model Capture and Curation | FAIR Lite 3: (Meta)Data | FAIR Lite 4: Storage |
|---|---|---|---|---|
| **Models should be Findable** | | | | |
| F1. Each model is assigned a globally unique and persistent identifier and different versions are assigned distinct identifiers. | **X** | | | |
| F2 Models are described with rich metadata covering all aspects of the model. (F2.1.) Models are associated with searchable metadata for the property or endpoint to be predicted. | | | **X** | |
| F2 Models are described with rich metadata covering all aspects of the model. (F2.2.) Models are associated with searchable metadata or descriptions of the chemicals (e.g. InCHI or SMILES), or chemical class(es), within the model, or a description of its applicability domain. | | **X** | **X** | |
| F3. Models' (meta)data clearly and explicitly include the identifier of the model they describe and are registered or indexed in a searchable resource. | **X** | | **X** | |
| F4. Models are registered or indexed in a searchable resource. (F4.1) Models' identifiers should be optimized to allow for use in multiple search engines. | **X** | | | **X** |
| **Models should be Accessible** | | | | |
| A1. Models are retrievable by their identifier using a standardized communications protocol. (A1.1.) The model (and any associated protocol represented by the model (meta)data) is openly accessible or re-implementable. | **X** | | | **X** |
| A1. Models are retrievable by their identifier using a standardized communications protocol. (A1.2.) The model (and any associated protocol) allows for an authentication and authorization procedure, where necessary. | **X** | | | **X** |
| A2. Model (meta)data are accessible even when the model is no longer available, unless restricted for commercial, ethical or data protection reasons (e.g., blinding of confidential chemical structures). | | | | **X** |
| **Models should be Interoperable** | | | | |
| I1. The models and their (meta)data are described in a standardized manner, i.e., standards to define chemical structures, endpoints, molecular descriptors and modelling algorithms. | | **X** | **X** | |
| I2. The model reads, writes and exchanges data in a way that meets domain-relevant community standards. | | **X** | | **X** |
| I3. The model must be interoperable with other software, e.g., with a clearly defined input/output i.e., with an appropriate Application Programming Interface (API) for shared web services. | | **X** | | **X** |
| I4. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. | | | **X** | |
| I5. (Meta)data use vocabularies that follow FAIR principles. | | **X** | **X** | **X** |
| I6. The model includes qualified references to other objects, such as molecular descriptors. | | **X** | | |
| **Models should be Reusable** | | | | |
| R1. The model is available for its use in some format (e.g., source code, executable, library or service). | | **X** | | **X** |
| R2. The usage license of the model should be clearly defined and appropriate to encourage its use. | | | **X** | |
| R3. The storage of the model and (meta)data should be done on a sustainable and future-proofed platform, anticipating the impact on the availability of software changes over time. | | **X** | | **X** |
| R4. Software includes qualified references to other software, e.g., so that the correct molecular descriptors can be obtained, either as part of the model or storage of the molecular descriptors software or experimental protocol. | | **X** | | **X** |
| R5. (Meta)data are richly described with a plurality of accurate and relevant attributes. (R5.1) The model and its (meta)data are associated with detailed provenance. | | **X** | **X** | |
| R6. The model and its (meta)data meet domain-relevant community standards for documentation. | | **X** | **X** | **X** |

*ALTEX, accepted manuscript*
*published September 25, 2025*
*doi:10.14573/altex.2502021*

**Tab. 3B: Mapping of the FAIR Principles for Research Software (FAIR4RS) (taken from Barker et al., 2022) against the FAIR Lite Principles**

| FAIR Principles for Research Software (FAIR4RS) (taken from Barker et al., 2022) | FAIR Lite 1: Unique Identifier | FAIR Lite 2: Model Capture and Curation | FAIR Lite 3: (Meta)Data | FAIR Lite 4: Storage |
|---|---|---|---|---|
| **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. | X | | | |
| F1. Software is assigned a globally unique and persistent identifier. F1.2. Different versions of the software are assigned distinct identifiers. | X | | | |
| F2. Software is described with rich metadata. | | | X | |
| F3. Metadata clearly and explicitly include the identifier of the software they describe. | X | | X | |
| F4. Metadata are FAIR, searchable and indexable. | | | X | X |
| **Software, and its metadata, is retrievable via standardized protocols** | | | | |
| A1. Software is retrievable by its identifier using a standardized communications protocol. A1.1. The protocol is open, free, and universally implementable. | X | X | | X |
| A1. Software is retrievable by its identifier using a standardized communications protocol. A1.2. The protocol allows for an authentication and authorization procedure, where necessary. | X | X | | X |
| A2. Metadata are accessible, even when the software is no longer available. | X | X | X | |
| **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. | | X | | X |
| I2. Software includes qualified references to other objects. | | X | | X |
| **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 license. | | | X | X |
| R1. Software is described with a plurality of accurate and relevant attributes. R1.2. Software is associated with detailed provenance. | | X | X | X |
| R2. Software includes qualified references to other software. | | X | | X |
| R3. Software meets domain-relevant community standards. | | X | | X |

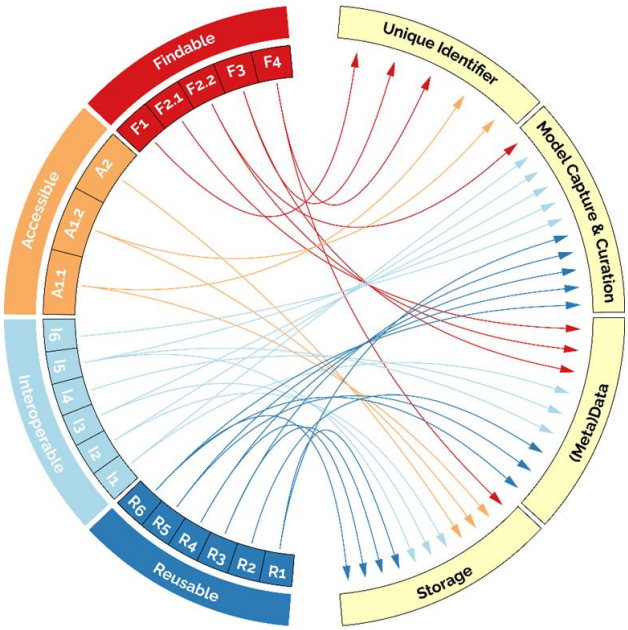


**Fig. 1: Relationships between FAIR principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) and FAIR Lite principles for *in silico* models**

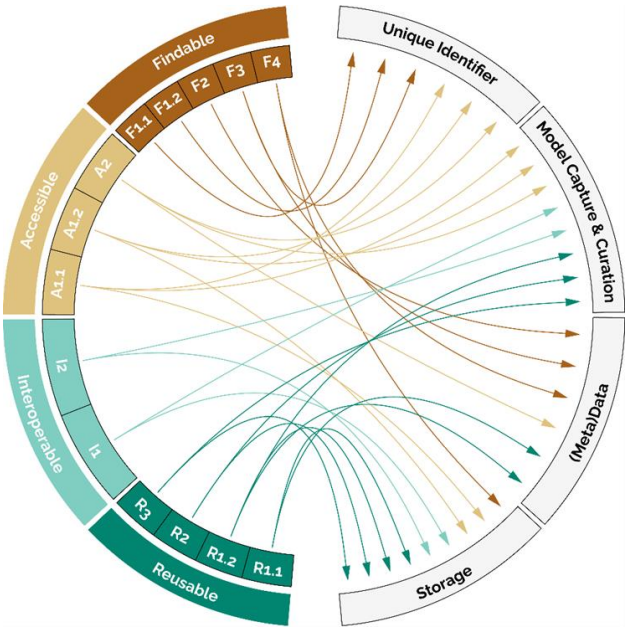


**Fig. 2: Relationships between FAIR4RS principles (Barker et al., 2022) and FAIR Lite principles for *in silico* models**

**Tab. 4: A checklist for the FAIR Lite criteria for computational toxicology models and solutions to ensure compliance**

| FAIR Lite Principle | Checklist | Solutions |
|---|---|---|
| **1. Unique identifier** | Is a unique and permanent identifier provided for the model? | A Document Object Identifier (DOI) is searchable and ubiquitous. |
| **2. Model capture and curation** | Is the model described and / or documented in a standardized ontology that meets community standards? Specifically for open models: is the model representation in a standardized format that can be downloaded and used? | The capture of the model should include the description of the training series and model engine (sufficient to allow the model to be reproduced), in addition to the parameters generated within the model. Where appropriate, the model should be represented in a machine-readable format, such as Predictive Model Markup Language (PMML), Open Neural Network Exchange (ONNX), etc., so it can be utilized. Standardized ontologies should be used to describe the model (see Table S1). Where standard ontologies are not available, terms from e.g. OECD Test Guidelines, regulatory guidance, current terminology in the scientific literature, etc., should be used. |
| | Is the applicability domain of the model stated? | A viable description of the applicability domain, encompassing chemical, property, metabolic, mechanistic information as appropriate. |
| | Are performance, diagnostic and quality assurance data and information provided? | As appropriate and required, estimates of model fit and predictivity should be available. Preferably, all model-related data should be available to calculate any relevant statistical metrics. |
| **3. (Meta)data** | Are the dependent (input) (meta)data of the model described and / or provided (for an open access model)? | Dependent data, e.g. toxicity values (and units), on which the model was developed should be described, i.e. number and type of chemicals, and, where appropriate, values provided. Actual values should be provided unless restricted by the license agreement or for commercial reasons. |
| | Are the independent (descriptor) (meta)data of the model described and / or provided (for an open access model)? | Independent data, e.g. physico-chemical properties, molecular descriptors for QSARs, coordinates, parameters for simulations etc. should be described, i.e. details on measurements, software version, if appropriate, and, where appropriate, values provided. Actual values should be provided unless restricted by the license agreement or for commercial reasons. |
| | Are (meta)data made available using standardized ontologies meeting community standards? | Standardized ontologies should be used to describe the dependent and independent data. Where standard ontologies are not available, terms from e.g. OECD Test Guidelines, regulatory guidance, etc. should be used. |
| | Is a license agreement available? | Restrictions on the use of the model, or an appropriate license agreement, e.g. Creative Commons, commercial, should be provided. |
| **4. Storage** | Is the information on the data, model, license, etc. stored on a suitable platform allowing for searching and retrieval? | A relevant storage platform such as QSAR Databank (QsarDB), BioModels etc. should be used. |
| | Is the code (or other software) of the model available for open access models or available from a third party e.g. a software vendor? | Code, or the model could be made available through a bespoke storage platform or via a public storage facility, e.g. GitHub, Zenodo etc. |
| | Can the model be integrated with other software to enable efficient use? | An API, or similar, is provided should the code or model not be immediately usable. |

QSARs (OECD, 2007). However, FAIR data and models are required for their assessment and validation, especially through automated means. To encourage the FAIRification of computational models, the questions in Table 4 (or adaptations thereof) could be implemented into procedures such as the QSAR Assessment Framework (QAF) (OECD, 2023; Gissi et al., 2024) or within *in silico* toxicology protocols (Myatt et al., 2018). The QAF will not, of course, be of use or appropriate for all of the broad range of computational models described in Tables 1 and 2. Therefore, it may also be appropriate to add in the requirement for the development of tiered strategies, e.g., DAs, IATAs, NGRA, etc., that the models used have a FAIR Lite checklist completed – this would enable a researcher to repeat the approach or apply it to another problem formulation.

The proposed FAIR Lite principles fit well into the FAIR principles as applied to QSARs (Cronin et al., 2023; Belfield et al., 2025) (Fig. 1) and other considerations already published. There is little doubt that FAIR principles, firstly proposed for data sharing (Wilkinson et al., 2016), crystalize the ethical and pragmatic considerations of all toxicologists and risk assessors using toxicology models and data. This is specifically with regard to making data and technologies, including the derivation of new knowledge based on the models, available and sharable for use. The FAIR Lite principles can also guide the development of federated systems where multiple QSARs models of various organizations are co-trained on proprietary data to optimize a single QSAR model without direct access to raw data. When all parties involved in collaborations adhere to FAIR principles in documenting their predictive models, it federated learning will foster greater trust and facilitates the adoption of privacy-enhancing methods (Smajić et al., 2023).

There have been a number of applications of the FAIR principles to areas within toxicology, notably to make new approach methodologies (NAMs) transparent and accessible (Colborne et al., 2023), with areas such as nanotoxicology taking the lead (Ammar et al., 2024; Barrick et al., 2023; Bossa et al., 2021; Furxhi et al., 2023; Jeliazkova et al., 2021) as well as being applied to Adverse Outcome Pathways (AOPs) (Mortensen et al., 2022; 2025; Wittwehr et al., 2024) and chemoinformatics technologies (Steinbeck, 2025). The current FAIR Lite principles aim to ensure that a user can find and

reuse a computational toxicology model to make predictions to support chemical safety assessment. This is particularly relevant to the plethora of published models currently not findable or usable, and to ensure that this does not occur again. It is intended that adherence to the FAIR Lite principles, as a minimum, should facilitate the process of better storage and searching for models.

Whilst the FAIR Lite checklist is intended to be straightforward for use, it is not intended to be a "pass/fail", but rather a qualitative indication of where a model may fall short of the overall FAIR principles and progress, or update is required. Pragmatism will undoubtedly be necessary to implement the principles, at least for the short-term. For instance, we do not currently have the technologies for standardizing toxicological terms and ontologies to comply fully with the "Model Capture and Curation" FAIR Lite principle. Despite the lack of an overarching ontology, the requirements have been identified and proposals made for such a framework (Hardy et al., 2012), as well as for QSARs (Russman et al., 2014; Sild et al., 2020; Spjuth et al., 2010) and nanomaterials (Hastings et al., 2015). A selection of existing, relevant ontologies is summarized in Table S1. The ontologies in Table S1 cover various aspects of the description of models and (meta)data. It is clear that to make these ontologies usable for computational toxicology effort is required to combine the relevant terms from them into the required overarching ontology. It should be recognized that resources are needed to support the FAIRification of computational models. Whilst FAIR is a laudable aim, it will not succeed without more significant input into the platforms for sustainable models and the development of globally accepted community standards to describe data and models.

## 4 Conclusions

A variety of computational toxicology models will be required to implement non-animal chemical safety assessments. The application of the FAIR principles to computational toxicology models is essential. This study has confirmed that certain fundamental criteria should be met to make a computational toxicology model FAIR. Four fundamental FAIR Lite principles for computational toxicology models, based on methodological foundations of computational toxicology and unambiguously understood by practitioners such as developers and end-users, are proposed. The aim in their development was that they are easy to use and understand, as well as being applicable to all types of computational models. As such, the FAIR Lite principles simplify the process of demonstrating the FAIRness of a computational model for toxicology. The FAIR Lite principles include the need for a globally unique identifier, the provision of data for dependent and independent variables, standardized and accepted vocabularies, documentation and a storage platform to make models interoperable and searchable. It is recommended that these criteria, or adaptations thereof, are integrated formally into the process for the validation of computational toxicology (e.g., into the QSAR Assessment Framework (QAF)). Meeting the criteria for FAIRness of any computational model will require informatics infrastructure (i.e., to provide access to the training data and model engine, preferably within a searchable database), establishing minimal levels of (meta)data (with particular reference to commercial models), as well as creating and confirming the overarching ontologies to describe all aspects of the models.

## References

Ammar, A., Evelo, C., and Willighagen, E. (2024). FAIR assessment of nanosafety data reusability with community standards. *Sci Data 11*, 503. doi:10.1038/s41597-024-03324-x

Baltazar, M.T., Cable, S., Carmichael, P. L. et al. (2020). A Next-Generation Risk Assessment case study for coumarin in cosmetic products. *Toxicol Sci 176*, 236-252. doi:10.1093/toxsci/kfaa048

Barber, C., Fowkes, A., Hanser T. et al. (2024). From model performance to decision support – The rise of computational toxicology in chemical safety assessments. *Comput Toxicol 31*, 100303. doi:10.1016/j.comtox.2024.100303

Barker, M., Chue Hong, N. P., Katz, D. S. et al. (2022). Introducing the FAIR Principles for research software. *Sci Data 9*, 622. doi:10.1038/s41597-022-01710-x

Barrick, A., Métais, I., Ettajani, H. P. et al. (2023). Establishing FAIR (Findable, Accessible, Interoperable and Reusable) principles for estuarine organisms exposed to engineered nanomaterials. *Int J Data Sci Analyt 16*, 407–419. doi:10.1007/s41060-023-00447-z

Belfield, S. J., Basiri, H., Chavan, S. et al. (2025). Moving towards making (quantitative) structure-activity relationships ((Q)SARs) for toxicity-related endpoints Findable, Accessible, Interoperable and Reusable (FAIR). *ALTEX* in press. doi:10.14573/altex.2411161

Bossa, C., Andreoli, C., Bakker, M. et al. (2021). FAIRification of nanosafety data to improve applicability of (Q)SAR approaches: A case study on *in vitro* Comet assay genotoxicity data. *Comput Toxicol 20*: 100190. doi:10.1016/j.comtox.2021.100190

Cherkasov, A., Muratov, E. N., Fourches, D. *et al.* (2014). QSAR modeling: Where have you been? Where are you going to? *J Med Chem 57*, 4977–5010. doi:10.1021/jm4004285

Colborne, J. K, and PrecisionTox Consortium. (2023). The Precision Toxicology initiative. *Toxicol Lett 383*, 33-42. doi:10.1016/j.toxlet.2023.05.004

Cronin, M. T. D., Belfield, S. J., Briggs, K. A. et al. (2023). Making *in silico* predictive models for toxicology FAIR. *Reg Toxicol Pharmacol 140*, 105385. doi:10.1016/j.yrtph.2023.105385

Cronin, M. T. D., Berggren, E., Camorani, S. et al. (2025). Report of the European Commission workshop on "*The Roadmap Towards Phasing Out Animal Testing for Chemical Safety Assessments*", Brussels, 11–12 December 2023. *Reg Toxicol Pharmacol 161*, 105818. doi:10.1016/j.yrtph.2025.105818

Delrue, N., Sachana, M., Sakuratani, Y. et al. (2016). The adverse outcome pathway concept: A basis for developing regulatory decision-making tools. *Altern Lab Anim 44*, 417-429. doi:10.1177/026119291604400504

De Vivo, M., Masetti, M., Bottegoni, G. et al. (2016). Role of molecular dynamics and related methods in drug discovery. *J Med Chem 59*, 4035–4061. doi:10.1021/acs.jmedchem.5b01684

Freudenthal, O., Da Silveira, M., & Deladiennee, L. (2024). Unlocking the potential of data harmonization and FAIRness in chemical risk assessment: lessons from practice and insights for policy development. *Environ Sci Eur 36*, 194. doi:10.1186/s12302-024-01022-4

Furxhi, I., Willighagen, E., Evelo, C. et al. (2023). A data reusability assessment in the nanosafety domain based on the NSDRA framework followed by an exploratory quantitative structure activity relationships (QSAR) modeling targeting cellular viability. *NanoImpact 31*, 100475. doi:10.1016/j.impact.2023.100475

Gissi, A., Tcheremenskaia, O., Bossa, C. et al. (2024). The OECD (Q)SAR Assessment Framework: A tool for increasing regulatory uptake of computational approaches. *Comput Toxicol 31*, 100326. doi:10.1016/j.comtox.2024.100326

Hardy, B., Apic, G., Carthew, P. et al. (2012). Toxicology ontology perspectives. *ALTEX 29*, 139-156. doi:10.14573/altex.2012.2.139

Hastings, J., Jeliazkova, N., Owen, G. et al. (2015) eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. *J Biomed Semant 6*, 10. doi:10.1186/s13326-015-0005-5

Igarashi, Y., Nakatsu, N., Yamashita, T. et al. (2015). Open TG-GATEs: a large-scale toxicogenomics database. *Nucl Acids Res 43(Database issue)*, D921-D927. doi:10.1093/nar/gku955

Jacobsen, A., de Miranda Azevedo, R., Juty, N. et al. (2020) FAIR Principles: Interpretations and implementation considerations. *Data Intell 2*, 10–29. doi:10.1162/dint_r_00024

Jeliazkova, N., Apostolova, M. D., Andreoli, C. et al. (2021). Towards FAIR nanosafety data. *Nat Nanotechnol 16*, 644–654. doi:10.1038/s41565-021-00911-6

Kostal, J. (2023). Making the case for quantum mechanics in predictive toxicology - Nearly 100 years too late? *Chem Res Toxicol 36*, 1444–1450. doi:10.1021/acs.chemrestox.3c00171

Luijten, M. (2024). The ASPIS-initiated Safety Profiling Algorithm (ASPA): setting the stage for Next-Generation Risk Assessment. *Toxicol Lett 399*, Supplement 2: S42. doi:10.1016/j.toxlet.2024.07.123

Macmillan, D. S. and Chilton, M. L. (2019). A defined approach for predicting skin sensitisation hazard and potency based on the guided integration of *in silico*, *in chemico* and *in vitro* data using exclusion criteria. *Reg Toxicol Pharmacol 101,* 35-47. doi:10.1016/j.yrtph.2018.11.001

Madden, J. C., Enoch, S. J., Paini, A. et al. (2020). A review of *in silico* tools as alternatives to animal testing: principles, resources and applications. *Altern Lab Anim 48*, 146-172. doi:10.1177/0261192920965977

Mortensen, H. M., Martens, M., Senn, J. et al. (2022). The AOP-DB RDF: Applying FAIR Principles to the semantic integration of AOP data using the Research Description Framework. *Front Toxicol 4*, 803983. doi:10.3389/ftox.2022.803983

Mortensen, H. M., Gromelski, M., Hench, G. et al. (2025). The FAIR AOP Roadmap for 2025: Advancing Findability, Accessibility, Interoperability, and Re-Usability of Adverse Outcome Pathways. *Comput Toxicol*, *in press*.

Myatt, G. J., Ahlberg, E., Akahori, Y. et al. (2018). *In silico* toxicology protocols. *Reg Toxicol Pharmacol 96*, 1-17. doi:10.1016/j.yrtph.2018.04.014

Nahal, Y., Menke, J., Martinelli, J. et al. (2024). Human-in-the-loop active learning for goal-oriented molecule generation. *J Cheminform 16*, 138. doi:10.1186/s13321-024-00924-y

OECD (Organisation for Economic Cooperation and Development) (2007). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships*. ENV/JM/MONO, vol 2, OECD, Paris, p. 154

OECD (Organisation for Economic Cooperation and Development) (2023). *(Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions*, OECD Series on Testing and Assessment, No. 386, OECD Publishing, Paris, doi:10.1787/d96118f6-en

Pastor, M., Quintana, J., Sanz, F. (2018) Development of an Infrastructure for the Prediction of Biological Endpoints in Industrial Environments. Lessons Learned at the eTOX Project. *Front Pharmacol 9*, doi:10.3389/fphar.2018.01147

Piir, G., Kahn, I., García-Sosa, A. T. et al. (2018). Best practices for QSAR model reporting: Physical and chemical properties, ecotoxicity, environmental fate, human health and toxicokinetics endpoints. *Environ Heal Persp 126*, 126001. doi:10.1289/EHP3264

Rowland, M., Peck, C., Tucker, G. (2011). Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu Rev Pharmacol Toxicol 51*, 45–73. doi:10.1146/annurev-pharmtox-010510-100540

Ruusmann, V., Sild, S., Maran U. (2014). QSAR DataBank - an approach for the digital organization and archiving of QSAR model information. *J Cheminf 6*, 25. doi:10.1186/1758-2946-6-25

Ruusmann, V., Sild, S., Maran U. (2015). QSAR DataBank repository: open and linked qualitative and quantitative structure–activity relationship models. *J Cheminf* 7, 32. doi:10.1186/s13321-015-0082-6

Safford, B., Api, A. M., Barratt, C. et al. (2015). Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. *Regul Toxicol Pharmacol 72*, 673–682. doi:10.1016/j.yrtph.2015.05.017

Smajić, A., Grandits, M., Ecker, G. F. (2023). Privacy-preserving techniques for decentralized and secure machine learning in drug discovery. *Drug Discov Today 28*, 103820. doi:10.1016/j.drudis.2023.103820

Sepehri, S., Heymans, A., De Win, D. et al. (2025). The TOXIN knowledge graph: supporting animal-free risk assessment of cosmetics. *Database 2025*, baae121. doi:10.1093/database/baae121

Sild, S., Piir, G., Neagu, D., Maran, U. (2020). Storing and using quantitative and qualitative structure–activity relationships in the era of toxicological and chemical data expansion, *in Big Data in Predictive Toxicology (series Issues in*

*Toxicology)*, Eds. D. Neagu, A. Richarz, Royal Society of Chemistry, 185-213. doi:10.1039/9781782623656-00185

Spînu, N., Cronin, M. T. D., Enoch, S. J. et al. (2020). Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Arch Toxicol 94*, 1497–1510. doi:10.1007/s00204-020-02774-7

Spjuth, O., Willighagen, E. L., Guha, R. et al. (2010). Towards interoperable and reproducible QSAR analyses: Exchange of datasets. *J Cheminf 2*, 5. doi:10.1186/1758-2946-2-5

Steinbeck, C. (2025). The evolution of open science in cheminformatics: a journey from closed systems to collaborative innovation. *J Cheminf 17*,: 44. doi:10.1186/s13321-025-00990-w

Sturla, S. J., Boobis, A. R., FitzGerald, R. E. et al. (2014). Systems toxicology: from basic research to risk assessment. *Chem Res Toxicol 27*, 314–329. doi:10.1021/tx400410s

Trisciuzzi, D., Alberga, D., Leonetti, F. et al. (2018). Molecular docking for predictive toxicology. In: Nicolotti, O. (ed) *Computational Toxicology*. *Methods in Molecular Biology 1800*. Humana Press, New York, NY. doi:10.1007/978-1-4939-7899-1_8

Wambaugh, J. F., Hughes, M. F., Ring, C. L. et al. (2018). Evaluating *in vitro–in vivo* extrapolation of toxicokinetics. *Toxicol Sci 163*, 152–169. doi:10.1093/toxsci/kfy020

Wang, Z., Chen, J. (2019). Background, tasks, modeling methods, and challenges for computational toxicology. In: Hong, H. (Ed.), *Advances in Computational Toxicology*. *Challenges and Advances in Computational Chemistry and Physics*, 30. Springer, Cham. doi:10.1007/978-3-030-16443-0_2

Westmoreland, C., Bender, H. J., Doe, J. E. et al. (2022). Use of New Approach Methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA Deep Dive Workshop. *Reg Toxicol Pharmacol 135*, 105261. doi:10.1016/j.yrtph.2022.105261

Wetmore, B. A., Wambaugh, J. F., Allen, B. et al. (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. *Toxicol Sci 148*, 121–136. doi:10.1093/toxsci/kfv171

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3*, 160018. doi:10.1038/sdata.2016.18

Williams, A. J., Grulke, C. M., Edwards, J. et al. (2017). The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J Cheminf 9*, 61. doi:10.1186/s13321-017-0247-6

Wittwehr, C., Clerbaux, L. A., Edwards, S. et al. (2024). Why adverse outcome pathways need to be FAIR. *ALTEX 41*, 50-56. doi:10.14573/altex.2307131

Yang, C., Cronin, M. T. D., Arvidson, K.B. et al. (2021). COSMOS next generation – A public knowledge base leveraging chemical and biological data to support the regulatory assessment of chemicals. *Comput Toxicol 19*, 100175. doi:10.1016/j.comtox.2021.100175

Yang, H., Lou, C., Li, W. et al. (2020). Computational approaches to identify structural alerts and their applications in environmental toxicology and drug discovery. *Chem Res Toxicol 33*, 1312–1322. doi:10.1021/acs.chemrestox.0c00006

## Disclaimer

The opinions expressed in this document reflect only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

## Conflict of interest

The QSAR Databank has been developed and is maintained by Prof Uko Maran at the Institute of Chemistry, University of Tartu, Tartu, Estonia.

## Data availability

No data were used in this investigation.

## Acknowledgements

This project receives funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 964537 (RISK-HUNT3R), and it is part of the ASPIS cluster; the Horizon Europe Framework Programme project Partnership for the Assessment of Risks from Chemicals (PARC, grant 101057014) and the QUANTUM-TOX - Revolutionizing Computational Toxicology with Electronic Structure Descriptors and Artificial Intelligence (QUANTUM-TOX) HORIZON-EIC-2023-PATHFINDEROPEN-01 Project number: 101130724. UM, SS, GP acknowledge support by the Ministry of Education and Research, Republic of Estonia, through Estonian Research Council (grant number PRG1509), Ministry of Climate, Republic of Estonia (grant 4-4/22/19), Ministry of Social Affairs, Republic of Estonia (grants 3-4/1593-1, 3-4/2332-1, 3-4/2668-1).