Simplifying the Standardised Registration of Scientific Simulation Workflows Using Easy-MODA

Professor Iseult Lynch

School of Geography, Earth and Environmental Sciences
University of Birmingham

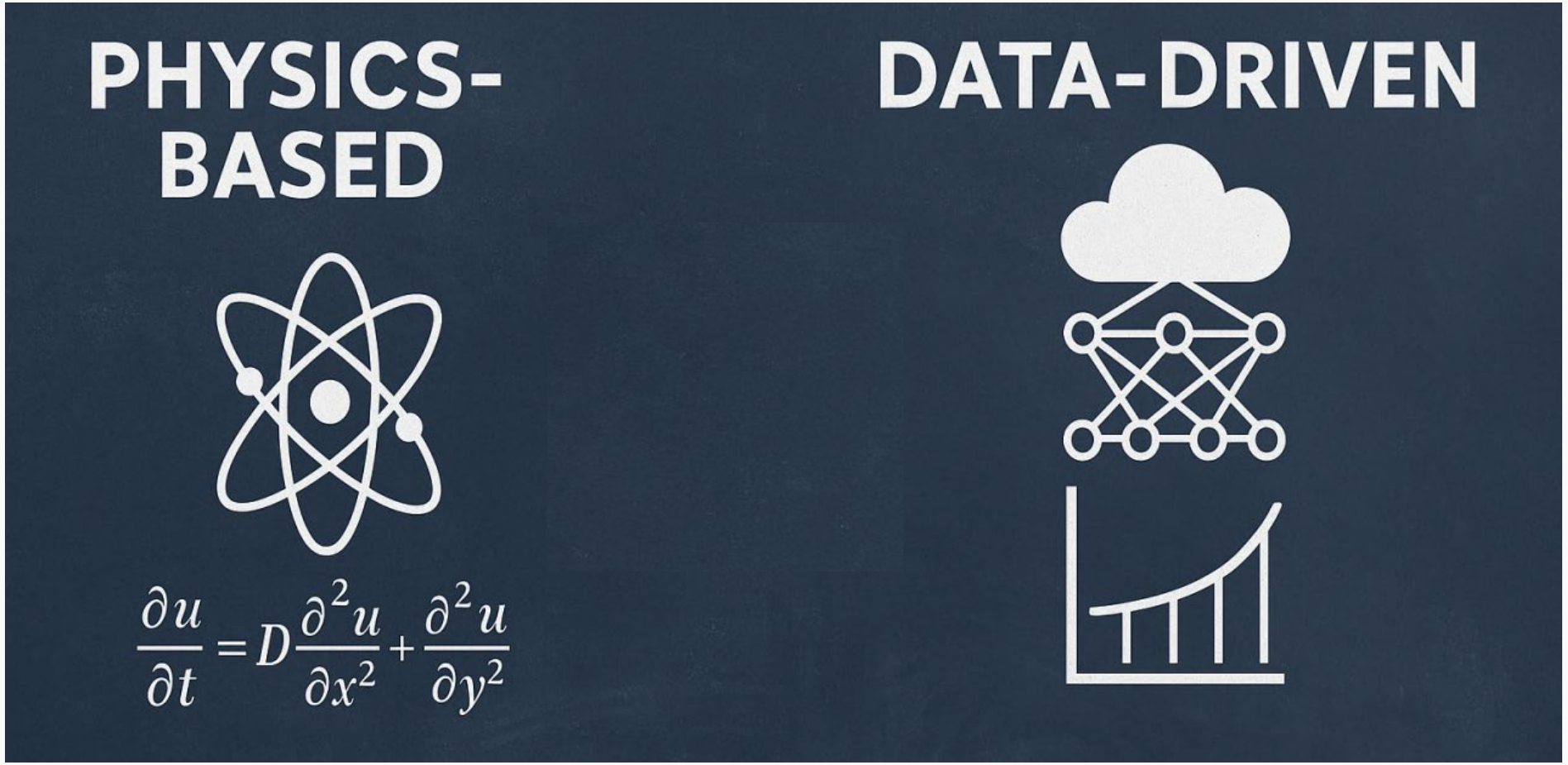
UNIVERSITY OF BIRMINGHAM

EMMC

INSIGHT

WorldFAIR

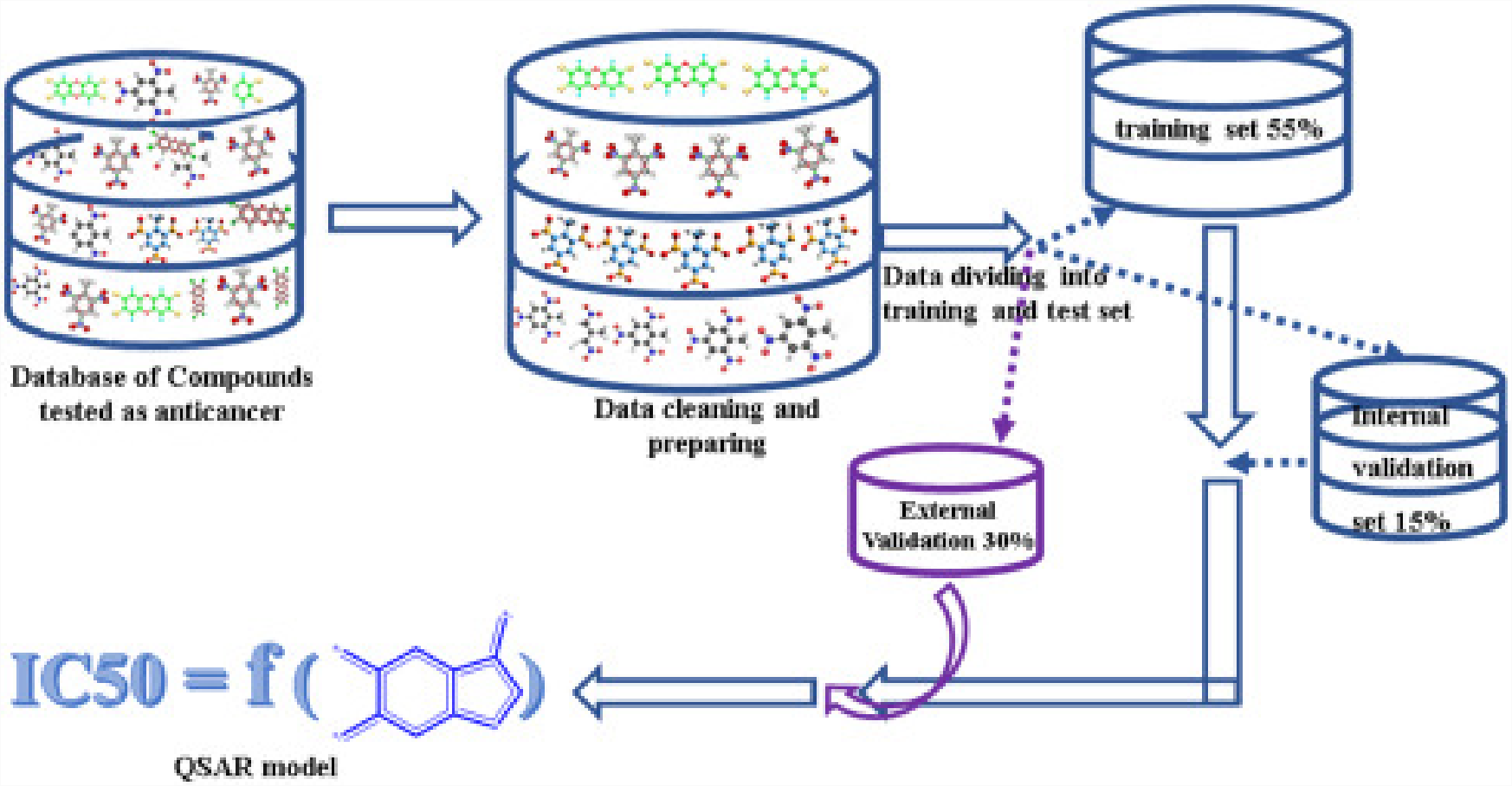
Physics-based versus data-driven models



UNIVERSITY OF BIRMINGHAM

Image from: https://www.linkedin.com/pulse/physics-based-vs-data-driven-modeling-opposites-allies-rahmanian-qsiof/

QSARs as the original data-driven models



UNIVERSITY OF BIRMINGHAM

https://doi.org/10.1016/B978-0-12-824156-1.00014-5

Reaching regulatory acceptance of QSARS took time!

The Acceptance of Computational Methods for the Regulatory Assessment of Chemicals

Andrew Worth

European Commission, Joint Research Centre, Institute for Health & Consumer Protection[1]

US Food & Drug Administration, Center for Food Safety & Applied Nutrition, Office of Food Additive Safety[2]

CIR workshop on SARs, Washington DC, 5 March 2012

[1] Permanent affiliation; [2] Current temporary affiliation

UNIVERSITY OF BIRMINGHAM

European Commission

https://www.cir-safety.org/sites/default/files/Worth_CIR_05-03-12%20[Compatibility%20Mode].pdf

The path to QSAR regulatory acceptance

Barriers to acceptance

8

- Industry assessors and regulators may not be familiar with QSAR methodology and therefore not comfortable using the results
- Modellers and regulators do not speak the same language
- Models might not be relevant to the regulatory question
- Models might not be sufficiently transparent
- Models might not be reproducible or readily available
- Insufficient practical guidance on how to use models in a regulatory context

UNIVERSITY OF BIRMINGHAM

https://www.cir-safety.org/sites/default/files/Worth_CIR_05-03-12%20[Compatibility%20Mode].pdf

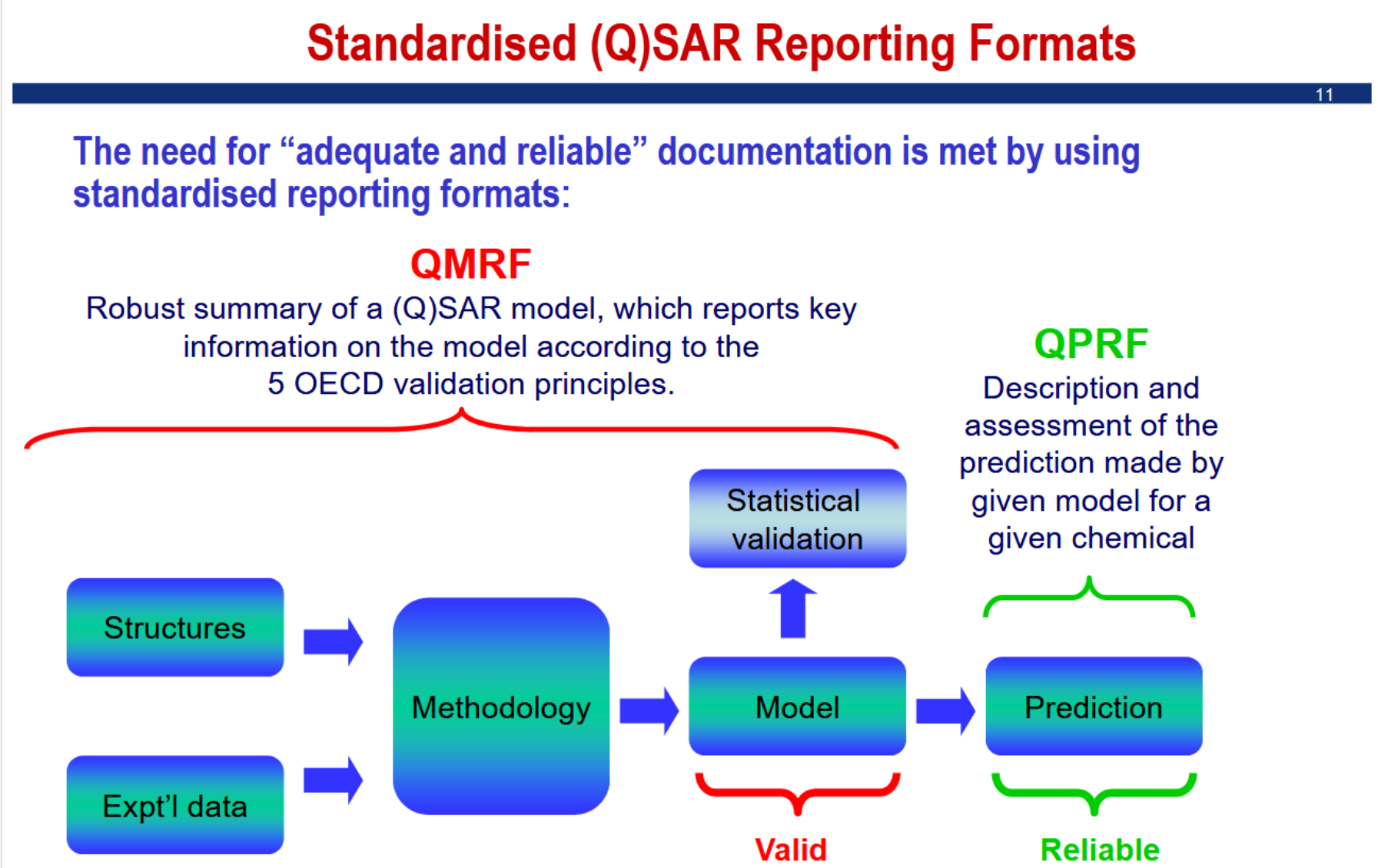
The path to QSAR regulatory acceptance

Overcoming the barriers to acceptance

- Consistent and internationally accepted reporting formats on QSAR models and their predictions: QMRF and QPRF
- Freely accessible software tools
- Training on how to use the tools
- Guidance on how to interpret the prediction results

https://www.cir-safety.org/sites/default/files/Worth_CIR_05-03-12%20[Compatibility%20Mode].pdf

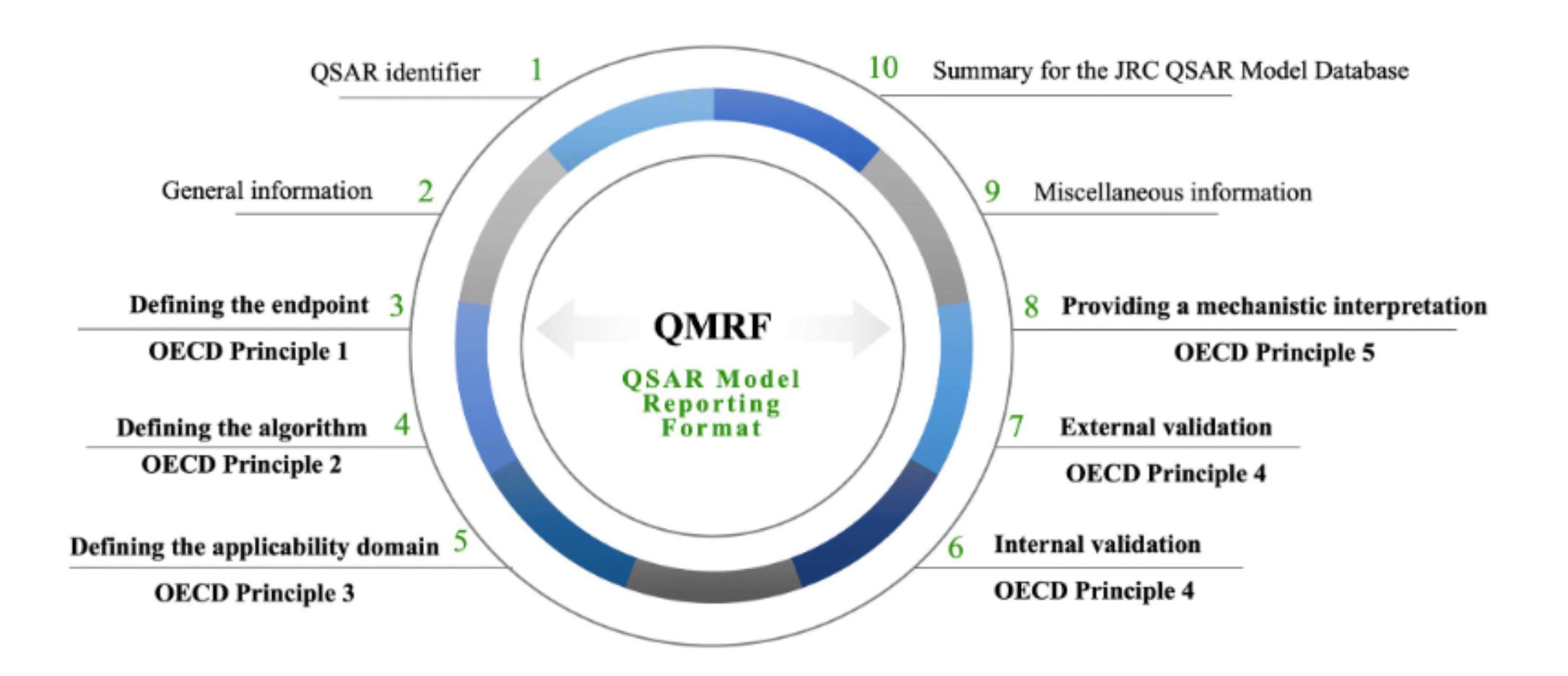
Model documentation key to regulatory acceptance!



UNIVERSITY OF BIRMINGHAM

https://www.cir-safety.org/sites/default/files/Worth_CIR_05-03-12%20[Compatibility%20Mode].pdf

QMRF for standardised model documentation (a FAIR Enabling Resource)



UNIVERSITY OF BIRMINGHAM

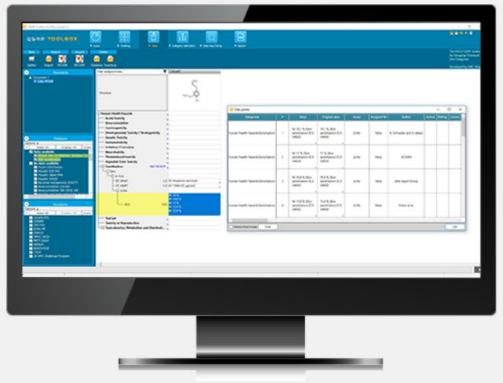
https://doi.org/10.5281/zenodo.10629631

QSAR modelling ecosystem

QSAR TOOLBOX

About Features Resources Support Developers Repository Download

MANUALS



Toolbox 4.7.1 Release Notes

DOWNLOAD PDF

Toolbox 4.7 Installation Manual

DOWNLOAD PDF

Toolbox 4.7 Multi-User Server Manual

DOWNLOAD PDF

Toolbox 4.7 Getting started – Quick reference guide

DOWNLOAD PDF

Deploying the TEDRA plugin

DOWNLOAD PDF

HELPDESK

UNIVERSITY OF BIRMINGHAM

QSAR modelling ecosystem

QSAR TOOLBOX

About Features Resources Support Developers Repository Download

About VEGA HUB

Biological, environmental and phys-chem properties of chemical substances are related to their structure

Using the VEGA platform, you can access a series of QSAR (quantitative structure-activity relationship) models for regulatory purposes, or develop your own model for research purposes.

QSAR models can be used to predict the property of a chemical compound, using information obtained from its structure.

The models have been taken from **CAESAR** or **T.E.S.T.**, or have been developed later by the contributors to **VEGA**.

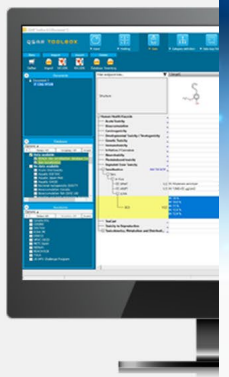
UNIVERSITY OF BIRMINGHAM

VEGA HUB QSAR Download VEGA HUB Community News Contacts

QSAR modelling ecosystem

QSAR TOOLBOX

About Features Resources Support Developers Repository Download



UNIVERSITY OF BIRMINGHAM

VEGA HUB

QMRFEditor

Download QMRFEditor 3.0.0

CHANGE LOG

- QMRFEditor-v3.0.0 (build date 31 Oct 2016)
- Updated QMRF DTD schema 3.0
- Ontology annotation (based on eNanoMapper ontology publication,ontology source)
- Free text search for ontology terms and protocols.
- Basic PMML support

Download QMRFEditor 2.0.0

CHANGE LOG

- QMRFEditor-v2.0.0 (build date 22 Jan 2013):
- replaced 'QMRF Inventory' with 'QMRF Database';
- updated the hint for QMRF section 10.1;
- the authors dropdown list is filled in from a remote QMRF inventory;
- by default saves into the QMRF XML only the endpoints used in the document;
- changed the DTD location to http://qmrf.sourceforge.net/qmrf.dtd

About

In the regulatory assessment of chem
Activity Models (QSAR) are playing a
for hazard and risk assessment. This
QSARs and to use them to derive est
estimates. To help meet these needs,
freely accessible through http://qsardb

The QSAR Model Reporting Format (
reporting key information on QSAR m
The information is structured accordin

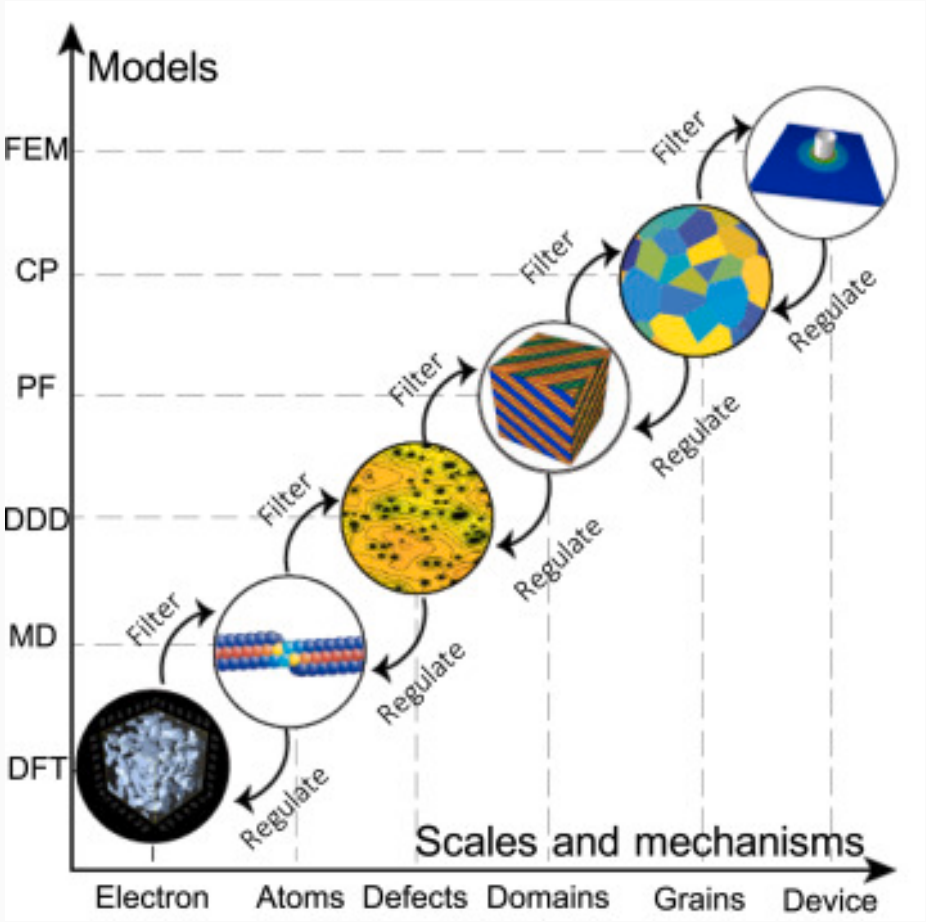
Download the QMRF Editor in order t
descriptions to JRC-COMPUTOX@ec

The QMRF schema 3.0 is updated Oc
enhancements applicable for reportin

Acknowledgements

> QMRF DTD schema, QMRF
> commissioned by JRC Com
> developed by Ideaconsult L
> QMRF 3.0.0 is updated with

Physics-based model documentation: MODA



PHYSICS-BASED MODEL

PHYSICS EQUATION
PE

Equation based on a **physical/chemical theory** which describes the spatial and temporal evolution of a chosen physics quantities of the entity

PHYSICS QUANTITIES

MATERIALS RELATIONS
MR

Information on the material needed to **close the PE** and to make the system of Governing Equations **solvable**

EXAMPLES

CLASSICAL MOLECULAR DYNAMICS

PE
Newton's equation of motion

$$\frac{dV}{dr} = -m\frac{d^2r}{dt^2}$$

MR
Lennard-Jones potential

$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

FLUID DYNAMICS

PE Navier Stokes equation

$$\frac{\partial}{\partial t}(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) = -\nabla\cdot p\mathbf{I} + \nabla\cdot\boldsymbol{\tau} + \rho\mathbf{g}$$

MR Stress tensor for incompressible flows

$$\nabla\cdot\boldsymbol{\tau} = 2\mu\nabla\cdot\boldsymbol{\varepsilon} = \mu\nabla\cdot\left(\nabla\mathbf{u} + \nabla\mathbf{u}^{\mathrm{T}}\right) = \mu\nabla^2\mathbf{u}$$

UNIVERSITY OF BIRMINGHAM

https://emmc.eu/moda/

Standardised documentation of Simulations (MODA)

The MODA comprises a **text template** and a **graphical workflow template** which should be used to document materials modelling workflows for EU projects, but they may also find use in organisations' documentations or in supplementary documents of scientific publications.

A compendium of over 100 projects and classification/terminology of materials modelling can be found in the **Review of Materials Modelling (RoMM).**

More information on MODA you may find in the presentation **"MODA - MOdelling DAta generalisation"** given within the EMMC 2017 International Workshop in Vienna.

UNIVERSITY OF BIRMINGHAM

https://emmc.eu/moda/

Standardised documentation of Simulations (MODA)

The MODA comprises a **text template** and a **graphical workflow template** which should be used to document materials modelling

upplementary documents of

be found in the **Review of Materials**

lisation"** given within the EMMC 2017

HOW IT LOOKS!

MODA for <user-case>
***Simulated* in project <acronym>**

OVERVIEW of the SIMULATION			
1	User Case		General description of the User Case. Please give the properties and behaviour of the particular material, manufacturing process and/or in-service-behaviour to be simulated. No information on the modelling should appear here. The idea is that this user-case can also be simulated by others with other models and that the results can then be compared.
2	Chain of Models	Model 1	Please identify the first model. Note these are assumed to be physics-based models unless it is specified differently. Most modelling projects consist of a chain of models, (workflow). Here only the Physics Equations should be given and only names appearing in the content list of the Review of Materials Modelling VI should be entered. This review is available on http://cc.europa.eu/research/industrial_technologies/e-library.cfm) All models should be identified as electronic, atomistic, mesoscopic or continuum.
		Model 2	Please identify the second model.
		Data-based model	If data-based models are used, please specify.
3	Publication Peer-Reviewing the data		Please give the publication which documents the data of this ONE simulation. This article should ensure the quality of this data set (and not only the quality of the models).
4	Access conditions		Please list whether the model and/or data are free, commercial or open source. Please list the owner and the name of the software or database (include a web link if available).
5	Workflow and its rationale		Please give a textual rationale of why you as a modeller have chosen these models and this workflow, knowing other modellers would simulate the same end-user case differently. This should include the reason why a particular aspect of the user case is to be simulated with a particular model.

1	Aspect of the User Case/System to be Simulated		
1.1	Aspect of the User Case to be simulated		Describe the aspects of the User Case textually. No modelling information should appear in this box. This case could also be simulated by other models in a benchmarking operation! The information in this chapter can be end-user information, measured data, library data etc. It will appear in the pink circle of your workflow picture. Simulated input which is calculated by another model should not be included (but this input is listed in chapter 2.4) Also the result of pre-processing necessary to translate the user case specifications to values for the physics variables of the entities can be documented here.
1.2	Material		Chemical composition, …
1.3	Geometry		Size, form, picture of the system (if applicable) Note that computational choices like simulation boxes are to be documented in chapter 3.
1.4	Time Lapse		Duration of the User Case to be simulated. This is the duration of the situation to be simulated. This is not the same as the computational times to be given in chapter 3.
1.5	Manufacturing process or in-service conditions		If relevant, please list the conditions to be simulated (if applicable). E.g. heated walls, external pressures and bending forces. Please note that these might appear as terms in the PE or as boundary and initial conditions, and this will be documented in the relevant chapters.
1.6	Publication on this data		Publication documenting the simulation with this single model and its data (if available and if not already included in the overall publication).

2	Generic Physics Of The Model Equation		
2.0	Model type and name		Model type and name **chosen from RoMM content list** (the PE). This PE and only this will appear in the blue circle of your workflow picture. Please do not insert any other text although an indication of the MR is allowed.
2.1	Model entity		The entity in this materials model is <finite volumes, grains, atoms, or electrons>
2.2	Model Physics/ Chemistry equation PE	Equation	Name, description and mathematical form of the PE In case of tightly coupled PEs set up as one matrix which is solved in one go, more than one PE can appear.
		Physical quantities	Please name the physics quantities in the PE, these are parameters (constants, matrices) and variables that appear in the PE, like wave function, Hamiltonian, spin, velocity, external force.
		Relation	Please, give the name of the Material Relation and which PE it completes.
2.3	Materials relations	Physical quantities/ descriptors for each MR	Please give the name of the physics quantities, parameters (constants, matrices) and variables that appear in the MR(s)
2.4	Simulated input		Please document the simulated input and with which model it is calculated. This box documents the interoperability of the models in case of sequential or iterative model workflows. Simulated output of the one model is input for the next model. Thus what you enter here in 2.4 will also appear in 4.1 of the model that calculated this input. If you do simulations in isolation, then this box will remain empty. Note that all measured input is documented in chapter 1 "User Case".

3	Solver and Computational translation of the specifications		
3.1	Numerical Solver		Please give name and type of the solver. E.g. Monte Carlo, SPH, FE, …iterative, multi-grid, adaptive,…
3.2	Software tool		Please give the name of the code and if this is your own code, please specify if it can be shared with an eventual link to a website/publication.
3.3	Time step		If applicable, please give the time step used in the solving operations. This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)
3.4	Computational Representation	Physics Equation, Material Relations, Material	Computational representation of the Physics Equation, Materials Relation and material. There is no need to repeat User Case info. "Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.
3.5	Computational boundary conditions		If applicable. Please note that these can be translations of the physical boundary conditions set in the User Case or they can be pure computational like e.g. a unit cell with mirror b.c. to simulate an infinite domain.
3.6	additional Solver Parameters		Please specify pure internal numerical solver details (if applicable), like • specific tolerances, • cut-off, convergence criteria • integrator options

4	POST PROCESSING		
4.1	The processed output		Please specify the output obtained by the post processing. If applicable then specify the entity in the next model in the chain for which this output is calculated: electrons, atoms, grains, larger/smaller finite volumes. In case of homogenisation, please specify the averaging volumes. Output can be calculated values for parameters, new MR and descriptor rules (data-based models).
4.2	Methodologies		Please describe the mathematics and/or physics used in this post-processing calculation. In homogenisation this is volume averaging. But also physics equations can be used to derive e.g. thermodynamics quantities or optical quantities from Quantum Mechanics raw output.
4.3	Margin Of Error		Please specify the margin of error (accuracy in percentages) of the property calculated and explain the reasons to an industrial end-user.

Standardised documentation of Simulations (MODA)

The MODA comprises a **text template** and a **graphical workflow template** which should be used to document materials modelling workflows for EU projects, but they may also find use in organisations' documentations or in supplementary documents of

HOW IT LOOKS!

MODA for <user-case>
***Simulated* in project <acronym>**

	OVERVIEW of the SIMULATION		
1	User Case		*General description of the User Case.* *Please give the properties and behaviour of the particular material, manufacturing process and/or in-service-behaviour to be simulated. No information on the modelling should appear here. The idea is that this user-case can also be simulated by others with other models and that the results can then be compared.*
2	Chain of Models	Model 1	*Please identify the first model. Note these are assumed to be physics-based models unless it is specified differently.* *Most modelling projects consist of a chain of models, (workflow). Here only the Physics Equations should be given and only names appearing in the content list of the Review of Materials Modelling VI should be entered. This review is available on http://ec.europa.eu/research/industrial_technologies/e-library.cfm) All models should be identified as electronic, atomistic, mesoscopic or continuum.*
		Model 2	*Please identify the second model.*
		data-based model	*If data-based models are used, please specify.*
3	Publication Peer-Reviewing the data		*Please give the publication which documents the data of this ONE simulation.* *This article should ensure the quality of this data set (and not only the quality of the models).*
4	Access conditions		*Please list whether the model and/or data are free, commercial or open source. Please list the owner and the name of the software or database (include a web link if available).*
5	Workflow and its rationale		*Please give a textual rationale of why you as a modeller have chosen these models and this workflow, knowing other modellers would simulate the same end-user case differently.* *This should include the reason why a particular aspect of the user case is to be simulated with a particular model.*

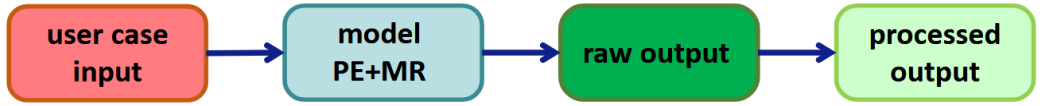
1	Aspect of the User Case/System to be Simulated	
1.1	Aspect of the User Case to be simulated	*Describe the aspects of the User Case textually.* *No modelling information should appear in this box. This case could also be simulated by other models in a benchmarking operation! The information in this chapter can be end-user information, measured data, library data etc. It will appear in the pink circle of your workflow picture.* *Simulated input which is calculated by another model should not be included (but this input is listed in chapter 2.4)* *Also the result of pre-processing necessary to translate the user case specifications to values for the physics variables of the entities can be documented here.*
1.2	Material	*Chemical composition, ...*
1.3	Geometry	*Size, form, picture of the system (if applicable)* *Note that computational choices like simulation boxes are to be documented in chapter 3.*
1.4	Time Lapse	*Duration of the User Case to be simulated.* *This is the duration of the situation to be simulated. This is not the same as the computational times to be given in chapter 3.*
1.5	Manufacturing process or in-service conditions	*If relevant, please list the conditions to be simulated (if applicable).* *E.g. heated walls, external pressures and bending forces. Please note that these might appear as terms in the PE or as boundary and initial conditions, and this will be documented in the relevant chapters.*
1.6	Publication on this data	*Publication documenting the simulation with this single model and its data (if available and if not already included in the overall publication).*

3	Solver and Computational Translation of the Specifications		
3.1	Numerical Solver		*Please give name and type of the solver.* *E.g. Monte Carlo, SPH, FE, ...Iterative, multi-grid, adaptive,...*
3.2	Software tool		*Please give the name of the code and if this is your own code, please specify if it can be shared with an eventual link to a website/publication.*
3.3	Time step		*If applicable, please give the time step used in the solving operations. This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)*
3.4	Computational Representation	Physics Equation, Material Relations, Material	*Computational representation of the Physics Equation, Materials Relation and material.* *There is no need to repeat User Case info. "Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.*
3.5	Computational boundary conditions		*If applicable.* *Please note that these can be translations of the physical boundary conditions set in the User Case or they can be pure computational like e.g. a unit cell with mirror b.c. to simulate an infinite domain.*
3.6	additional Solver		*Please specify pure internal numerical solver details (if applicable), like* • *specific tolerances,* • *cut-off, convergence criteria*

2	Generic Physics Of The Model Equation	
	Model type and name	*Model type and name* ***chosen from RoMM content list*** *(the PE).*

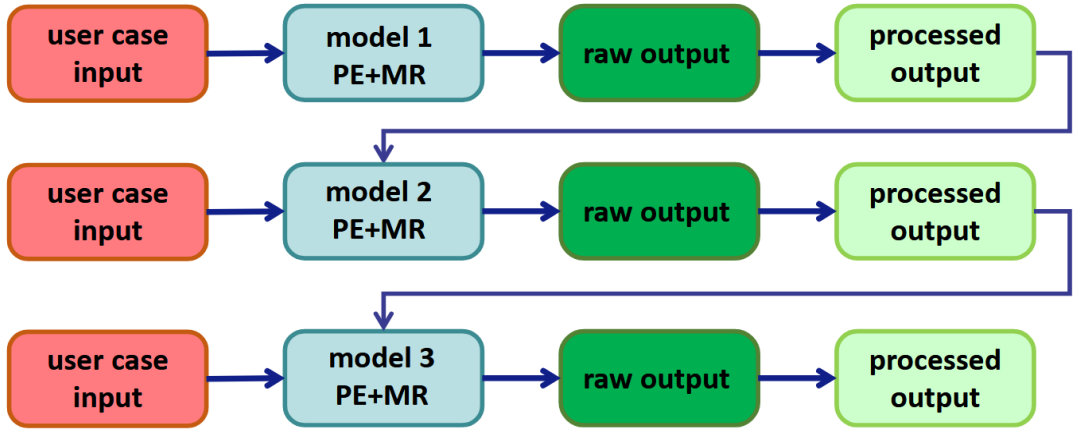
4.3	Please specify the margin of error (accuracy in percentages) of the property calculated and explain the reasons to an industrial end-user.

workflow for a **stand-alone model**

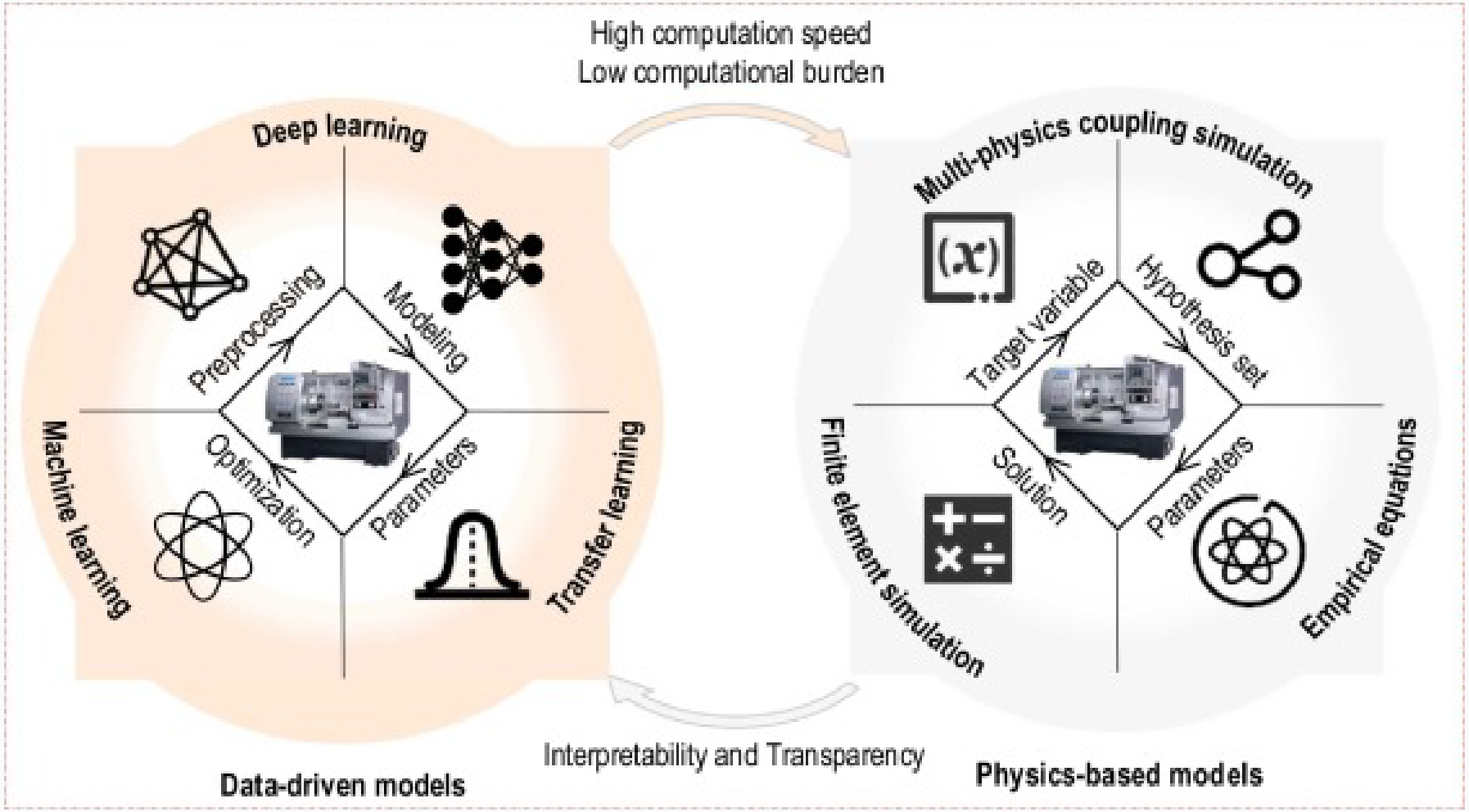


workflow for a chain of **linked models**

equations solved sequentially (i.e. one-way dependency)



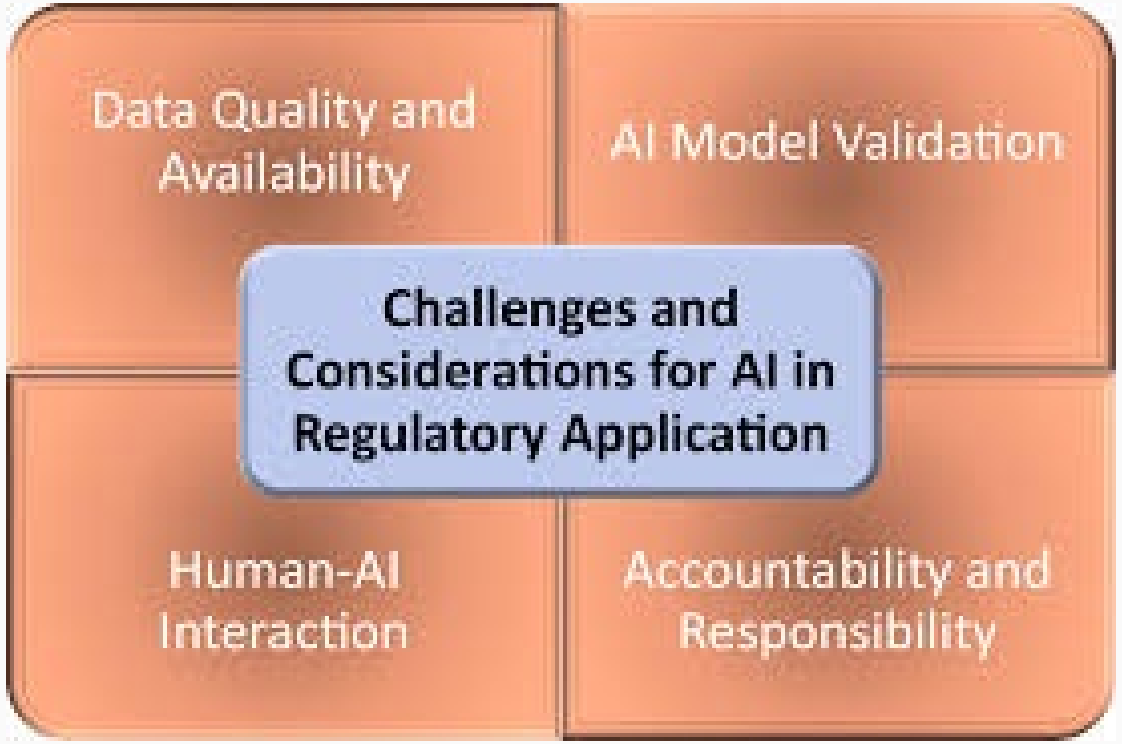
Reality is more complex: hybrid physics-data models

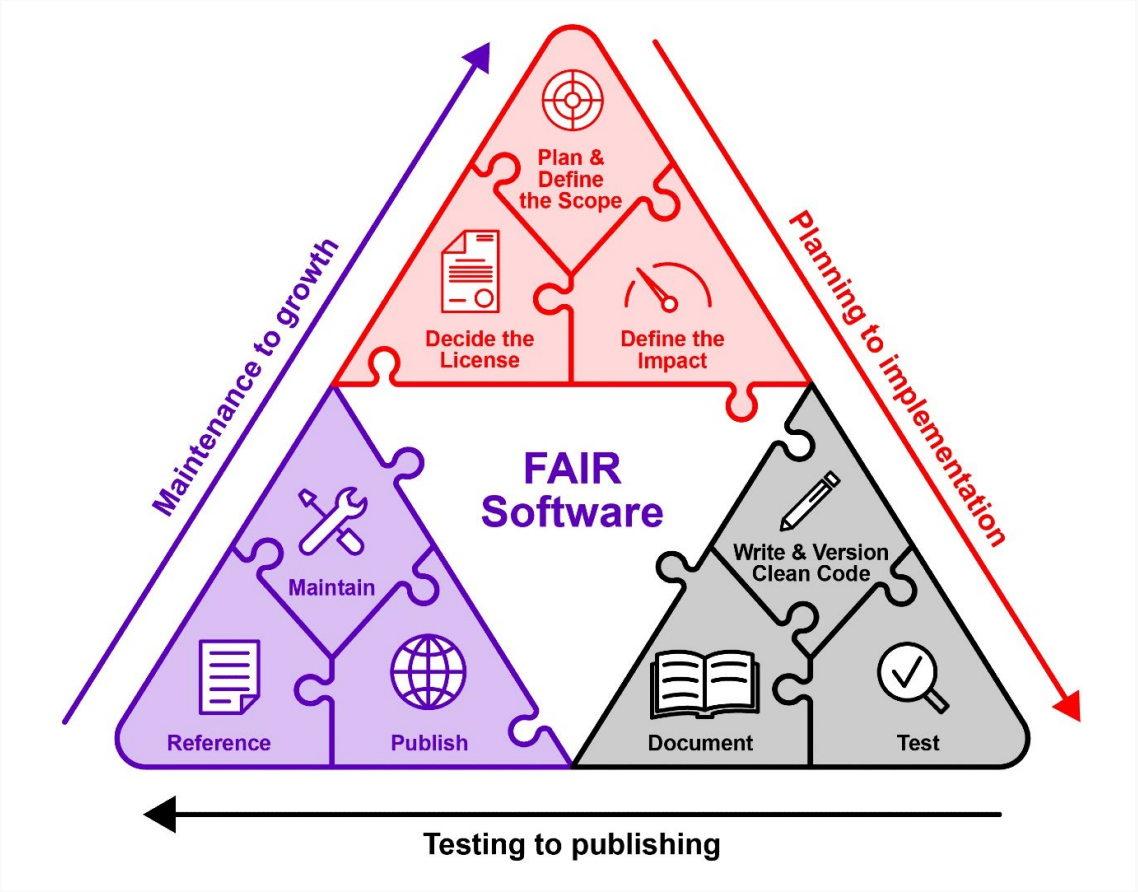


UNIVERSITY OF BIRMINGHAM

https://doi.org/10.1016/j.jmsy.2022.04.004

The drivers for model documentation – Regulatory acceptance and FAIR for models (software)



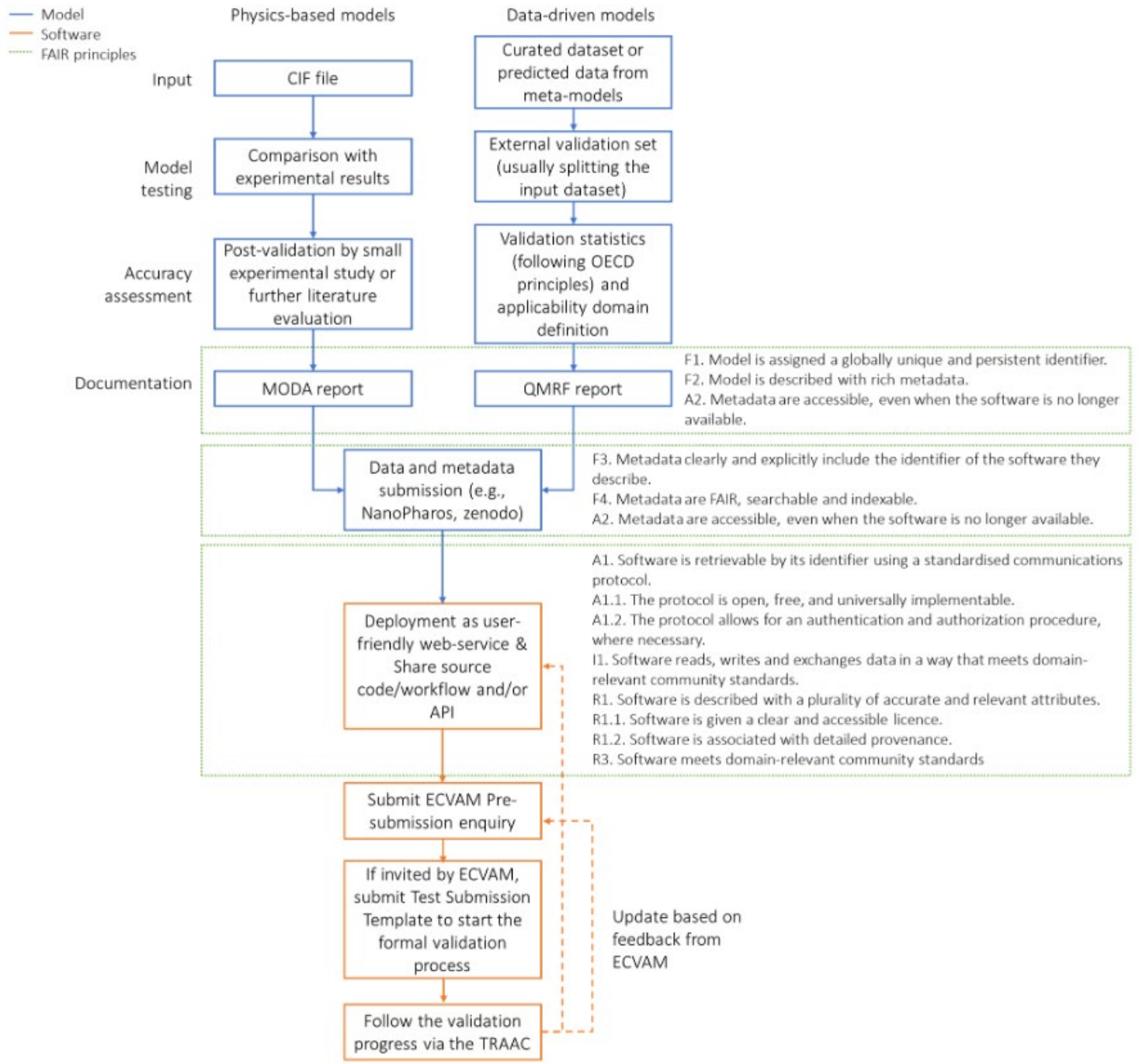


UNIVERSITY OF BIRMINGHAM

Lynch, I., & Afantitis, A. (2024). WorldFAIR (D4.2) FAIRification of nanoinformatics tools and models recommendations. Zenodo. https://doi.org/10.5281/zenodo.10629631

MODA supports FAIR models
(as QMRF does for QSAR models)

Physics-based models / Multi-modal models
need ecosystem of tools & services
=> EasyMODA



UNIVERSITY OF BIRMINGHAM

Tooling to facilitate model documentation

https://www.enaloscloud.novamechanics.com/insight/moda/

INSIGHT Easy-MODA WorldFAIR

Simplifying Standardized Registration of Scientific Simulation Workflows through MODA Template Guidelines powered by the Enalos Cloud Platform

Green

User Guide Load version from: Local or Cloud Search

MODA: MODELLING DATA GENERALISATION

Please fill in the boxes

Short title of the project

Enter the project title...

Acronym of the project

Enter the project acronym...

Description of the project

Enter the description of the project...

Tooling to reduce errors / formalise metadata and ontologies etc. (FAIRification)

Panagiotis (Takis) Kolokathis, NovaMechanics

Computational and Structural Biotechnology Journal 25 (2024) 256–268

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Software/web server article

Easy-MODA: Simplifying standardised registration of scientific simulation workflows through MODA template guidelines powered by the Enalos Cloud Platform

Panagiotis D. Kolokathis [a,b], Nikolaos K. Sidiropoulos [a,b], Dimitrios Zouraris [b,c], Dimitra-Danai Varsou [a,b], Dimitris G. Mintis [b,c], Andreas Tsoumanis [a,c], Francesco Dondero [b,d], Thomas E. Exner [e], Haralambos Sarimveis [f], Evgenia Chaideftou [g], Martin Paparella [g], Fotini Nikiforou [h], Achilleas Karakoltzidis [h], Spyros Karakitsios [h], Dimosthenis Sarigiannis [h], Jesper Friis [i,j], Gerhard Goldbeck [j,k], David A. Winkler [l,m,n], Willie Peijnenburg [o,p], Angela Serra [q], Dario Greco [q], Georgia Melagraki [r], Iseult Lynch [b,s,*], Antreas Afantitis [a,b,c,**]

UNIVERSITY OF BIRMINGHAM

EMMC

INSIGHT

WorldFAIR