

FAIR assessment of biosimulation models – a cross-community project

Irina Balaur^{1*}, Danielle Welter¹, David Nickerson², Rahuman Sheriff³, Matthias König⁴, Henning Hermjakob³, Reinhard Schneider¹, Venkata Satagopam¹ and Dagmar Waltemath^{5*}

1. Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg
2. Auckland Bioengineering Institute, University of Auckland, New Zealand
3. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), UK
4. Institute for Biology, Institute for Theoretical Biology, Humboldt University of Berlin, Germany
5. Medical Informatics Laboratory, University Medicine Greifswald, Germany

*Corresponding authors: irina.balaur@uni.lu and dagmar.waltemath@uni-greifswald.de

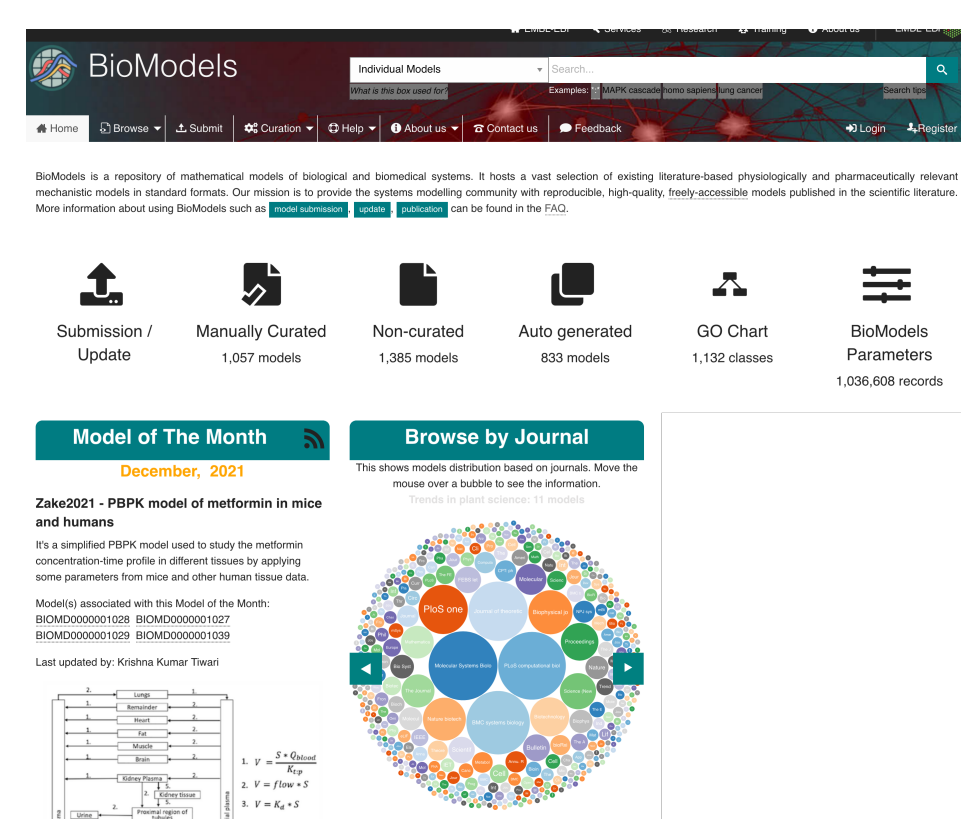
Background

Biosimulation Models

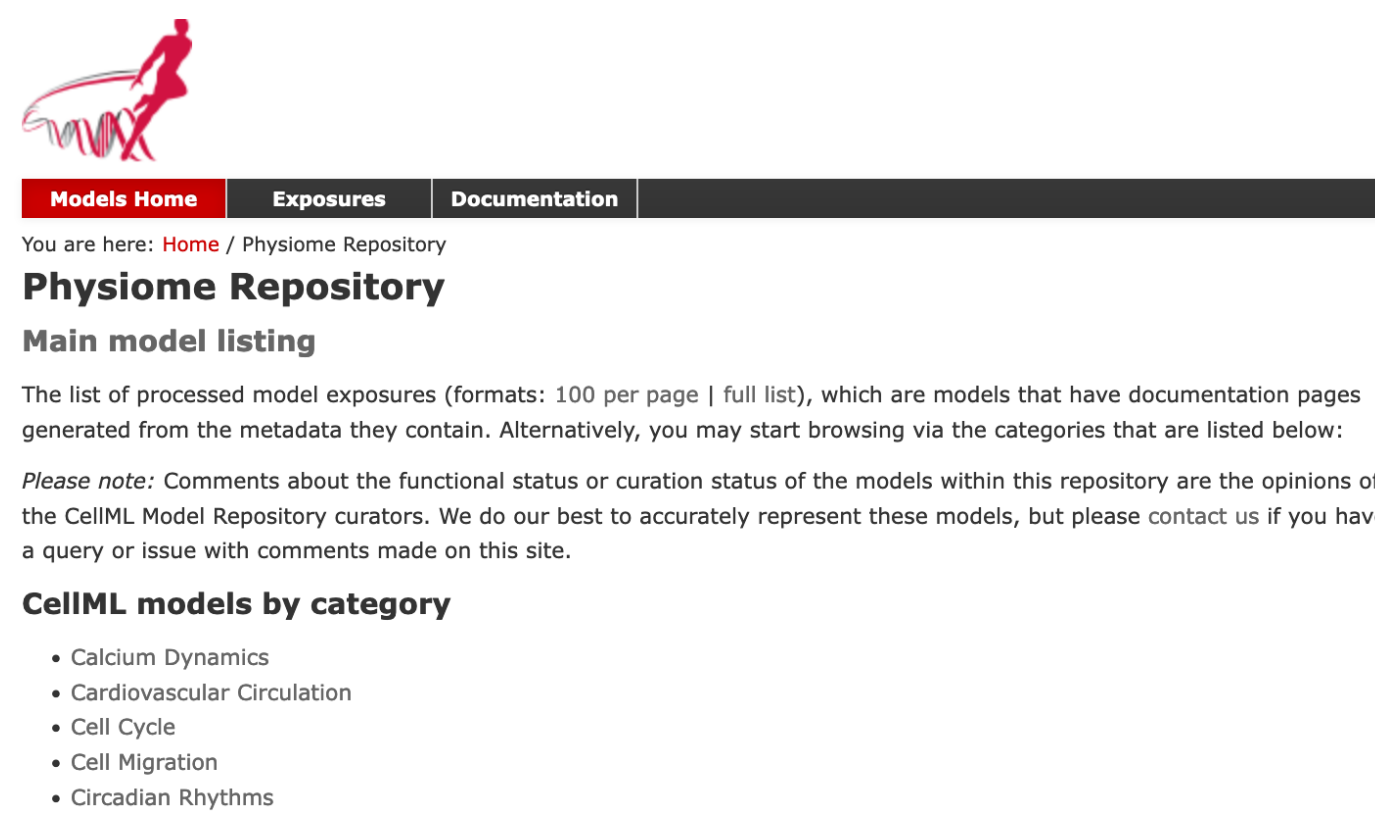
- Considerable ongoing development of biosimulation models
- Incorporating multi-layered information to explore complex disease mechanisms
- Potential as supporting tools for diagnosis and therapy

Model sharing

- Representation, visualisation and simulation model encoding in community-agreed standard formats
- Standardised curation- and publication processes, including BioModels and the Physiome Model Repository



<https://www.ebi.ac.uk/biomodels/>



<https://models.physiomeproject.org/>

Related work

Reproducibility Scorecard

- Assess the reproducibility of model-based results
- Total of eight questions for self-evaluation
- Score of 4 or higher indicates reproducibility

RDA FAIR data maturity Indicators

- “Lingua franca” to make results of FAIR evaluations comparable
- Generic model can be adapted to domain-specific needs
- Increased coherence and interoperability of FAIR assessment frameworks

Reproducibility Scorecard

The aim of this scorecard is to help evaluate the reproducibility of systems biology model and its simulation. It consists of eight questions with a unit score for each 'yes' as an answer. The higher the total score, the more reproducible the model is expected to be. All eight questions may not always be applicable and hence, on the scale of 8, a minimal score of 4 is recommended. Model authors, reviewers and journal editors can assess each systems biology model in a manuscript using this scorecard to support the peer-review process. For details see Twant et al., 2021 Mol Sys Bio <https://doi.org/10.15252/msb.20209982> and <https://www.ebi.ac.uk/biomodels/reproducibility>

Manuscript Title:		Reviewer:
Journal:		
Manuscript Identifier/ID:		Total Reproducibility Score:
Please check the tick box for a 'yes' answer to the questions below		
1. Are the mathematical expressions provided and explained in the manuscript/supplementary material?	☐	Yes
2. Are the parameters and entity initial levels in the manuscript/supplementary material (e.g., listed as a table)?	☐	Yes
3. Are simulation conditions including software/programming environment, algorithm, changes in parameters/concentration/states and any data normalization described under each simulation figure or attached as a supplementary material?	☐	Yes

<https://doi.org/10.15252/msb.20209982>

Table 1 FAIR data maturity model indicators.				
FAIR	ID	Indicator		Priority
F1	RDA-F1-01M	Metadata is identified by a persistent identifier	☐	Essential
F1	RDA-F1-01D	Data is identified by a persistent identifier	☐	Essential
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	☐	Essential
F1	RDA-F1-02D	Data is identified by a globally unique identifier	☐	Essential

<https://doi.org/10.5334/dsj-2020-041>

Preliminary results: FAIR evaluation

- A key step towards trust building and cross-discipline communication in relation to computational models is taken with the FAIR maturity model.
- FAIR can be a connecting principle across the clinical and biomedical domain as it is recognised and appreciated in both fields.

Development of FAIR model indicators

Sub-principle	ID	Indicator	Priority	Assessment overall	Assessment Essential	Assessment non-essential
Findable	F1	RDA-F1-01M	Metadata is identified by a persistent identifier	Essential	1	-
Findable	F1	RDA-F1-01D	Data is identified by a persistent identifier	Essential	1	-
Findable	F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	Essential	1	-
Findable	F1	RDA-F1-02D	Data is identified by a globally unique identifier	Essential	1	-
Findable	F2	RDA-F2-01M	Rich metadata is provided to allow discovery	Essential	1	-
Findable	F3	RDA-F3-01M	Metadata includes the identifier for the data	Essential	1	-
Findable	F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	Essential	1	-
Accessible	A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	Important	1	-
Accessible	A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	Essential	1	-
Accessible	A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	Essential	1	-
Accessible	A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	Essential	1	-
Accessible	A1	RDA-A1-03D	Data identifier resolves to a data object	Essential	1	-
Accessible	A1	RDA-A1-04M	Metadata is accessed through standardised protocol	Essential	1	-
Accessible	A1	RDA-A1-04D	Data is accessed through standardised protocol	Essential	1	-
Accessible	A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	Important	1	-
Accessible	A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	Essential	1	-
Accessible	A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	Important	1	-
Accessible	A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	Useful	1	-
Accessible	A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	Essential	1	-
Interoperable	I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	Important	0	-
Interoperable	I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	Important	1	-
Interoperable	I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	Important	1	-
Interoperable	I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	Important	1	-
Interoperable	I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	Important	0	-
Interoperable	I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	Useful	0	-
Interoperable	I3	RDA-I3-01M	Metadata includes references to other metadata	Important	1	-
Interoperable	I3	RDA-I3-01D	Data includes references to other data	Useful	0	-
Interoperable	I3	RDA-I3-02M	Metadata includes references to other data	Useful	1	-
Interoperable	I3	RDA-I3-02D	Data includes qualified references to other data	Useful	0	-
Interoperable	I3	RDA-I3-03M	Metadata includes qualified references to other data	Useful	1	-
Interoperable	I3	RDA-I3-04M	Metadata includes qualified references to other data	Useful	1	-
Reusable	R1	RDA-R1-01M	Plurality of accurate and relevant attributes are	Sum	30	18
Reusable	R1.1	RDA-R1.1-01M	Metadata includes information about the license			
Reusable	R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence			
Reusable	R1.1	RDA-R1.1-03M	Metadata refers to a machine-understandable			
Reusable	R1.2	RDA-R1.2-01M	Metadata includes provenance information acc.			
Reusable	R1.2	RDA-R1.2-02M	Metadata includes provenance information acc.			
Reusable	R1.3	RDA-R1.3-01M	Metadata complies with a community standard			
Reusable	R1.3	RDA-R1.3-01D	Data complies with a community standard			
Reusable	R1.3	RDA-R1.3-02M	Metadata is expressed in compliance with a m			
Reusable	R1.3	RDA-R1.3-02D	Data is expressed in compliance with a machi			

Score total = # fulfilled / # total (in relevant category)

Score applicable = # fulfilled / (# total - # NA) (in relevant category)

Legend –
indicator
priority:
Essential
Important
Useful

Priority	Assessment overall	Assessment Essential	Assessment non-essential
Sum	30	18	12
Score total	73.10%	90%	57.10%
Score applicable	73.10%	90%	57.10%

- Adapted from RDA indicators (DOI: 10.15497/rda00045)
- Proposal for a community-level standardized FAIR evaluation
- Binary scoring to calculate a level of FAIRness
- Application on model repositories, particularly in BioModels (<https://www.ebi.ac.uk/biomodels/>) and the Physiome Model Repository (<https://models.physiomeproject.org>)

Project goals

“Fostering the uptake of RDA indicators in Systems Biomedicine as a measure for model quality and FAIRness within the COMBINE community”

Deliverables:

1. FAIR model indicators with accompanying guidelines
2. Semi-automatic FAIR evaluation tool for use by the community

Vision:

- Implement FAIR evaluation as a standard procedure during model curation
- Provide a binary score for FAIRness level of individual models

Funding:

RDA/EOSC Future Cross-disciplinary Science Adoption (October 2022 – March 2023)



RESEARCH DATA ALLIANCE
EUROPE



Acknowledgments

Participants and collaborators at COMBINE 2021 and HARMONY 2021 for the FAIR model sessions. IB acknowledges the IMI FAIRplus Fellowship (no. 802750). IB and DaW acknowledge funding from the EOSC.