

Case study on FAIRification and certification: QsarDB

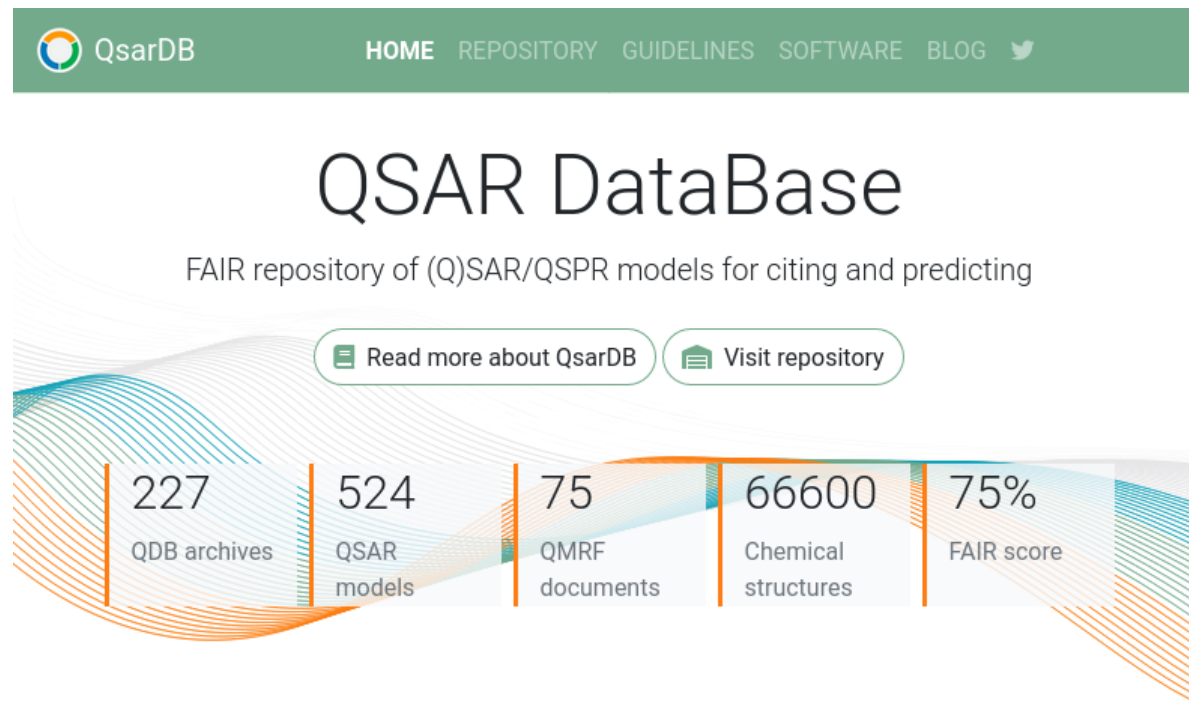
Sulev Sild

Institute of Chemistry

University of Tartu

Overview

- About QsarDB
- FAIRification
- CoreTrustSeal certification



Transparent QSARs

Create and upload QDB archive



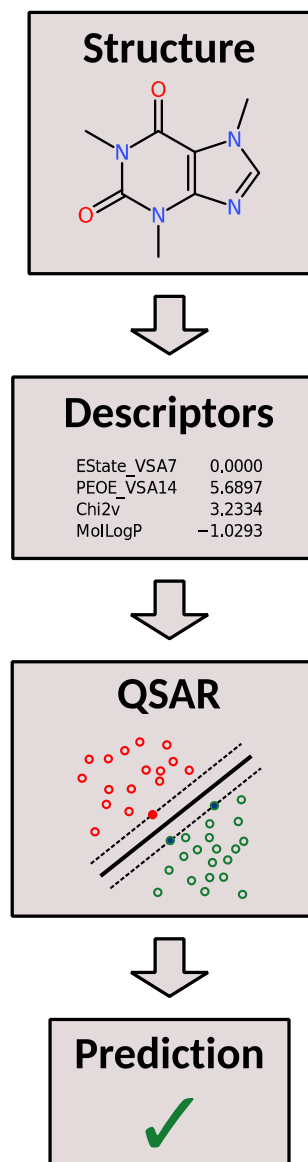
Cite QSAR data

Link repository with your publication



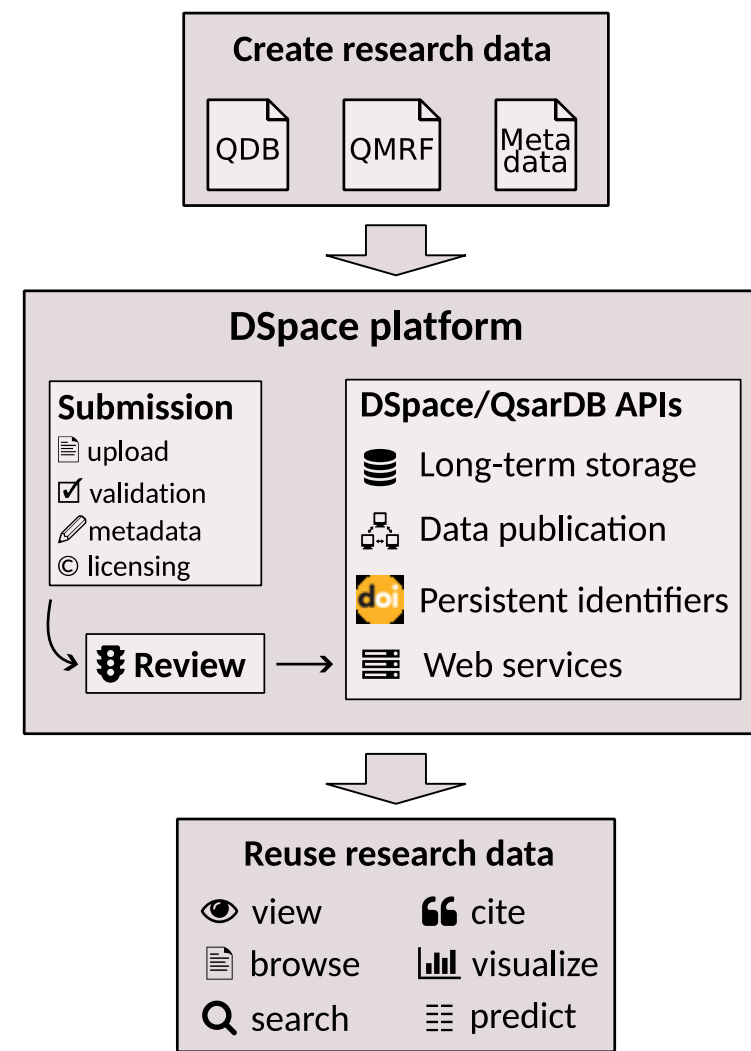
About QsarDB

- Open access data repository for (Q)SARs models and related data
- The repository is hosted at University of Tartu (since 2012)
- The main aim is to make (Q)SAR models transparent and reusable
 - Publishing and long-term archiving
 - Citable data publications for models (DOIs)
 - Cross-link model data with journal publications
 - Web services for using deposited models/data



QsarDB details

- Data policy
 - Prefer models from scientific publications
 - Accept data in QDB archive or QMRF formats
 - Open access
- QsarDB is based on modified DSpace 6.3
 - Automated validation in data submissions
 - Visualization of models and data
 - Descriptor calculation and prediction
 - Chemical search
- Data submissions are reviewed to ensure quality



Repository contents

227

QDB archives

75

QMRF documents

524

QSAR models

66600

Chemical structures

86 biological/chemical properties

- Phys. chem.
- Environmental fate
- Ecotoxic effects
- Human health
- Toxicokinetics
- Other

Model types

- Artificial Neural Network
- Counter-propagation Neural Network
- Decision Tree
- Ensemble Models
- k-Nearest Neighbors
- Logistic Regression
- Multi-Linear regression
- Random Forest
- Support Vector Machine

FAIR maturity before EOSC-Nordic collab.

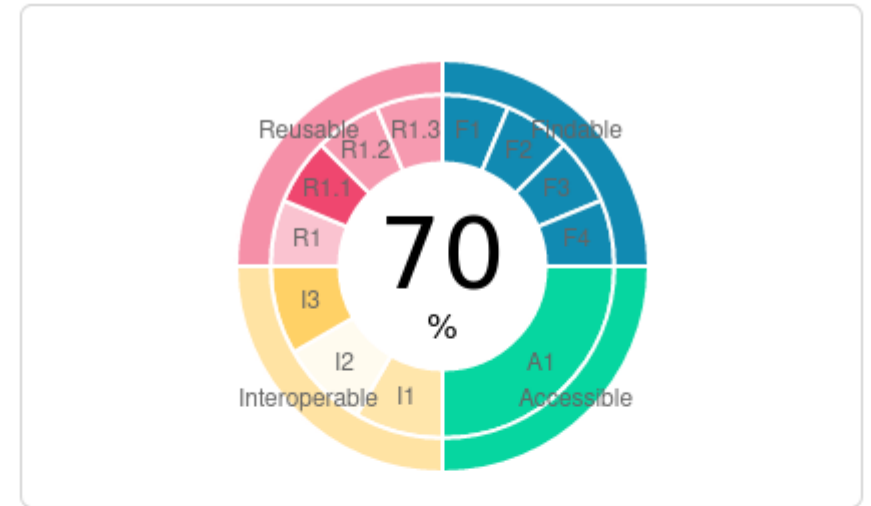
- First learned about the EOSC-Nordic project in 2020
- FAIR test results from F-uji tool for QsarDB (Dec 2020)

Name	Datasets	Platform	F-score	A-score	I-score	R-score	FAIR
QsarDB	20		42.86%	60.00%	50.00%	20.00%	36.67%

- May 2021 started collaboration with WP4
 - Analysed F-uji test results
 - Identified some low hanging fruit changes for improving the score

FAIRification results

- F-uji tool was very helpful
 - Very detailed output
 - Good support
- Reviewed metadata in QsarDB
 - Metadata was already in good state
 - Its machine readability was a problem
 - Improved consistency
 - Improved metadata in DataCite
- Hardest part was to configure/modify DSpace to make metadata machine readable
- We still have some room for improvement!



	Score earned:	Fair level:
Findable:	7 of 7 <input type="radio"/>	advanced
Accessible:	3 of 3 <input type="radio"/>	advanced
Interoperable:	2 of 4 <input type="radio"/>	moderate
Reusable:	5 of 10 <input type="radio"/>	moderate

CoreTrustSeal data repository certification

- We started the process started in 2021
 - Prepared the self-assessment document
 - Improved documentation on the homepage
 - Documented procedures
- EOSC-Nordic provided supported via
 - Zoom meetings
 - Reviewed the self-assessment document
- Documents were submitted to CoreTrustSeal this year
 - Received feedback
 - Some minor changes are needed
 - Resubmission is planned for this autumn



Summary

- Our collaboration with EOSC-Nordic has been valuable and productive
- With some help and a small development effort we managed to significantly improve FAIR maturity level in the QsarDB repository
- CoreTrustSeal certification process was much more difficult and it's moving on

Thank you!



Acknowledgements

- Uko Maran (QsarDB)
- Geven Piir (QsarDB)
- Robert Huber (PANGEA)
- Andreas Jaunsen (NordForsk)
- Mari Kleemola (Tampere University)
- Mari-Elisa Kuusniemi (Helsinki University)