UPCI
Unit of Process Control & Informatics,
National Technical University Of Athens, Greece

NanoCommons

Nano-Knowledge Community

Online Jaqpot Hackathon

Take your research from the bench to the community
by making your models
available as a web service

Philip Doganis, Pantelis Karatzas, Irene Liampa, Dimitra-Danai Varsou, Periklis Tsiros, Haralambos Sarimveis

NanoCommons Workshop, 2020-06-03, #nanocommons

This project has received funding from the European Union Horizon 2020 Programm (H2020) under grant agreement no. 731032.

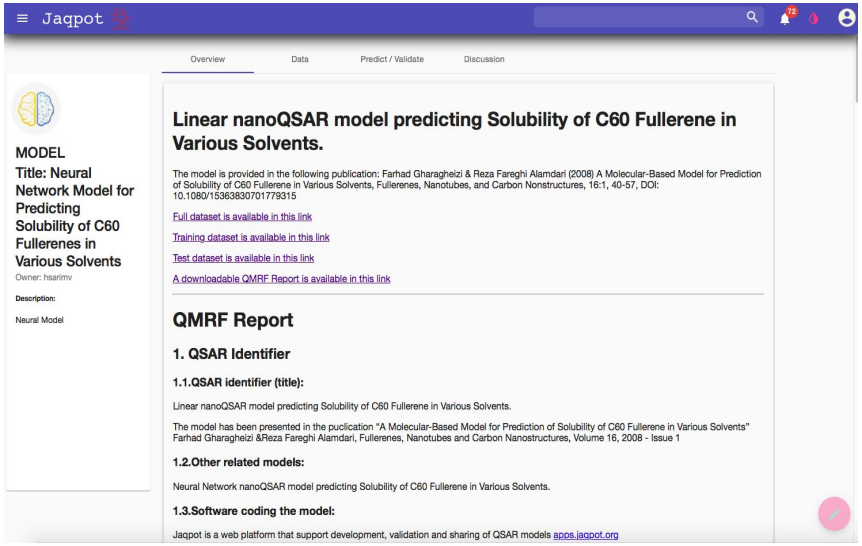
A place to

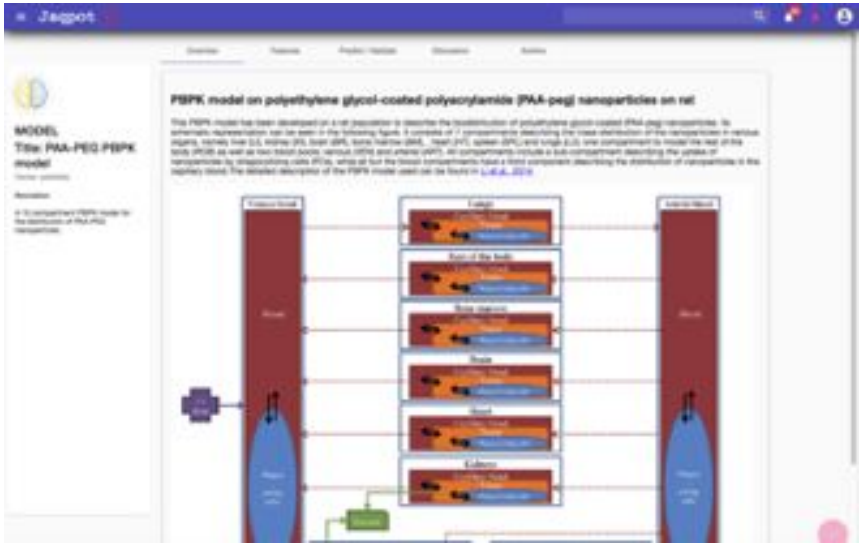
create discuss

share use

your models & datasets

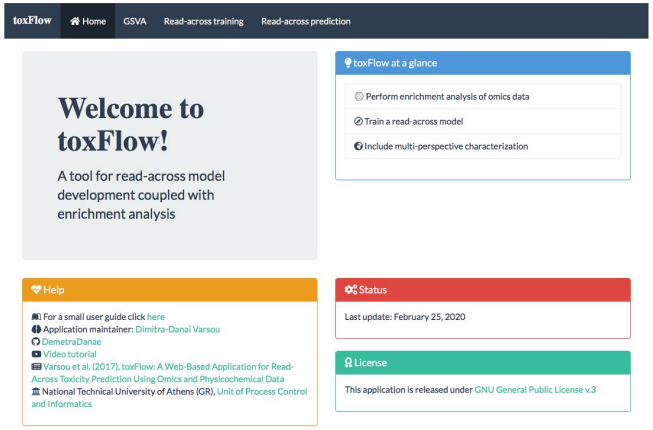
Jaqpot: a suite of tools for QSAR, PBPK, image analysis, read-across methods





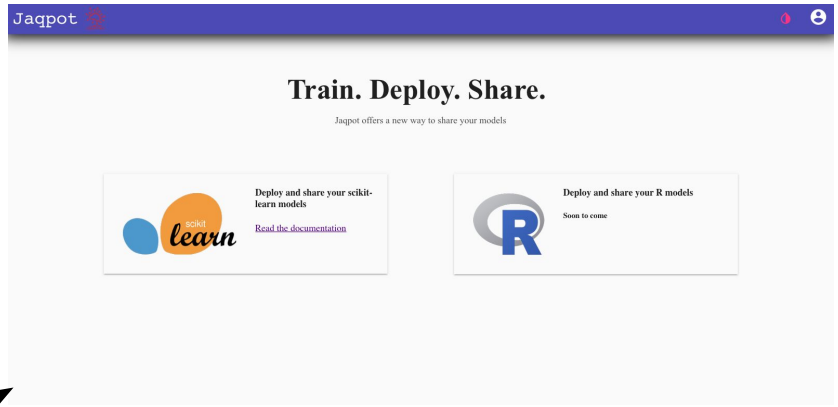
https://app.jaqpot.org/

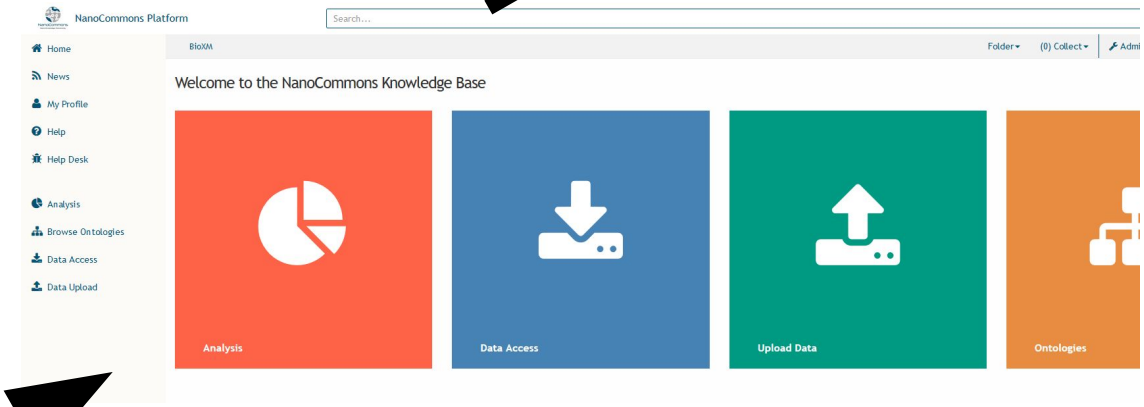
https://nanoimage.jaqpot.org



toxflow.jaqpot.org

NanoCommons KnowledgeBase & Jaqpot





Original Databases

NanoMILE/NanoFASE/NanoReg/eNanoMapper

Infrastructure

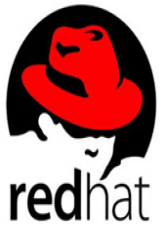
Docker is used to run software packages called "containers". Containers are isolated from each other and bundle their own application, tools, libraries a nd configuration files;

An open-source system for automating deployment, scaling, and management of containerized applications

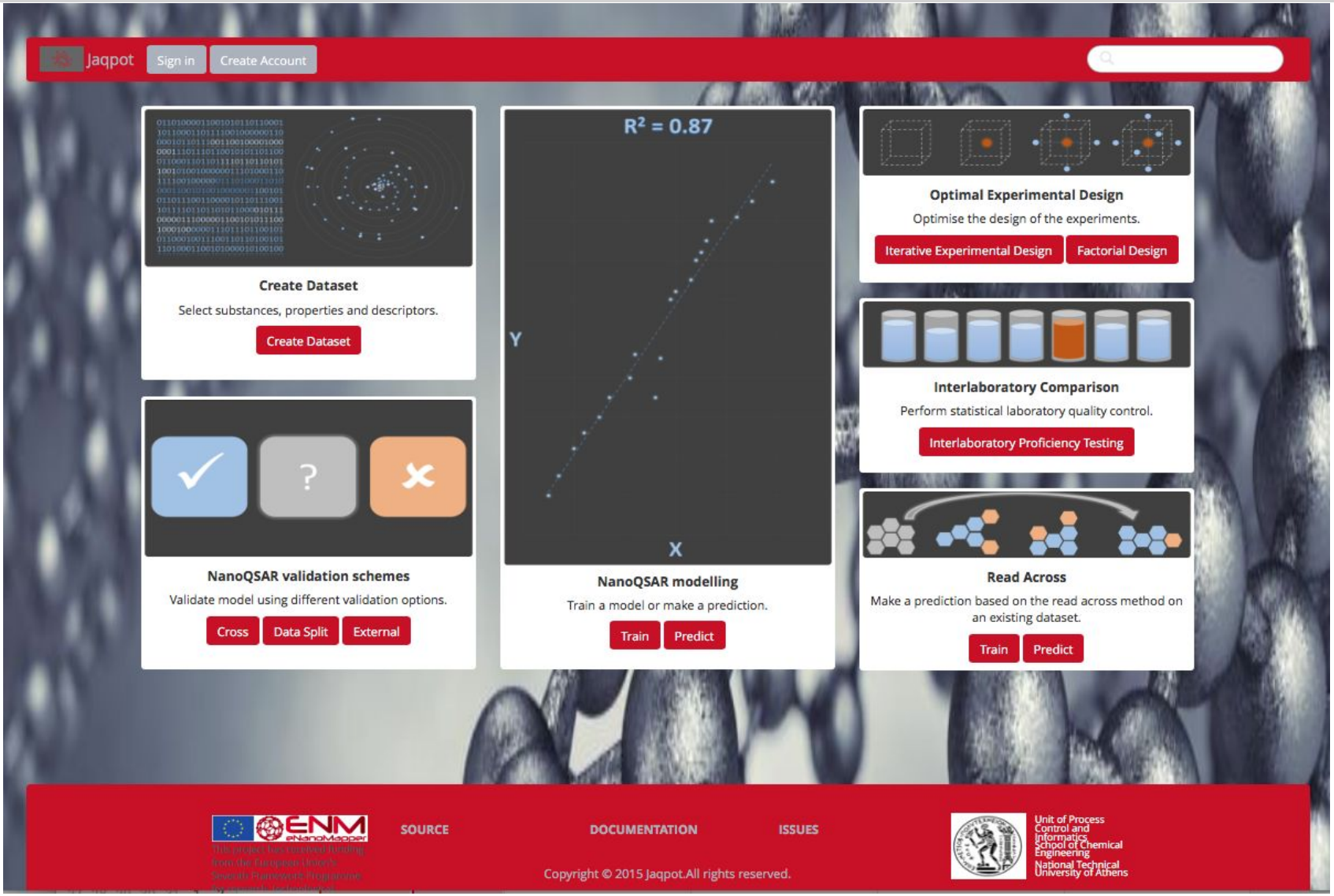
OpenShift is a multifaceted, open source container application platform from Red Hat Inc. for the development, deployment and management of applications.

An open source software product to allow OpenID Connect single sign-on with Identity Management and Access Management aimed at modern applications and services.

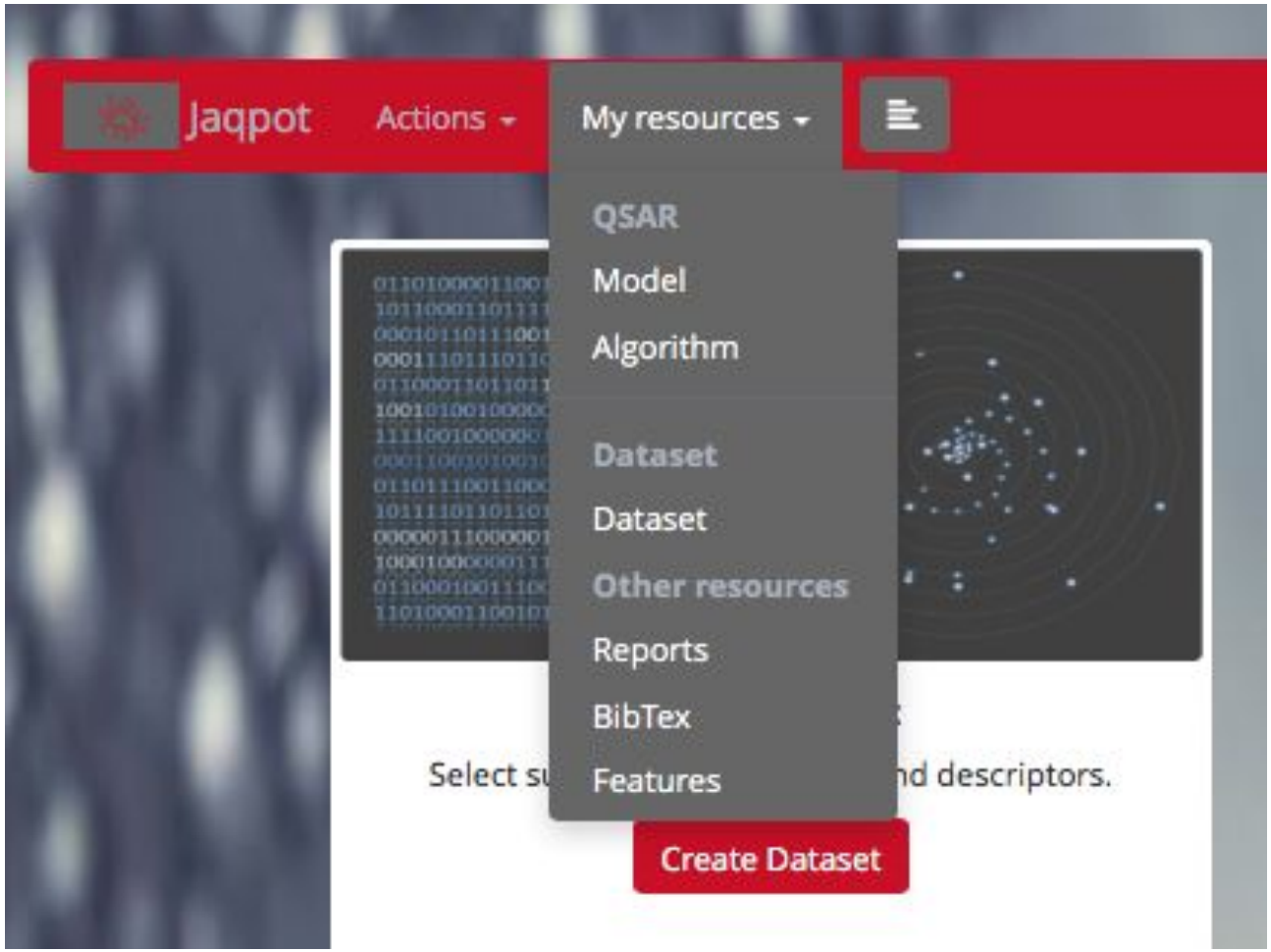
Under the stewardship of:



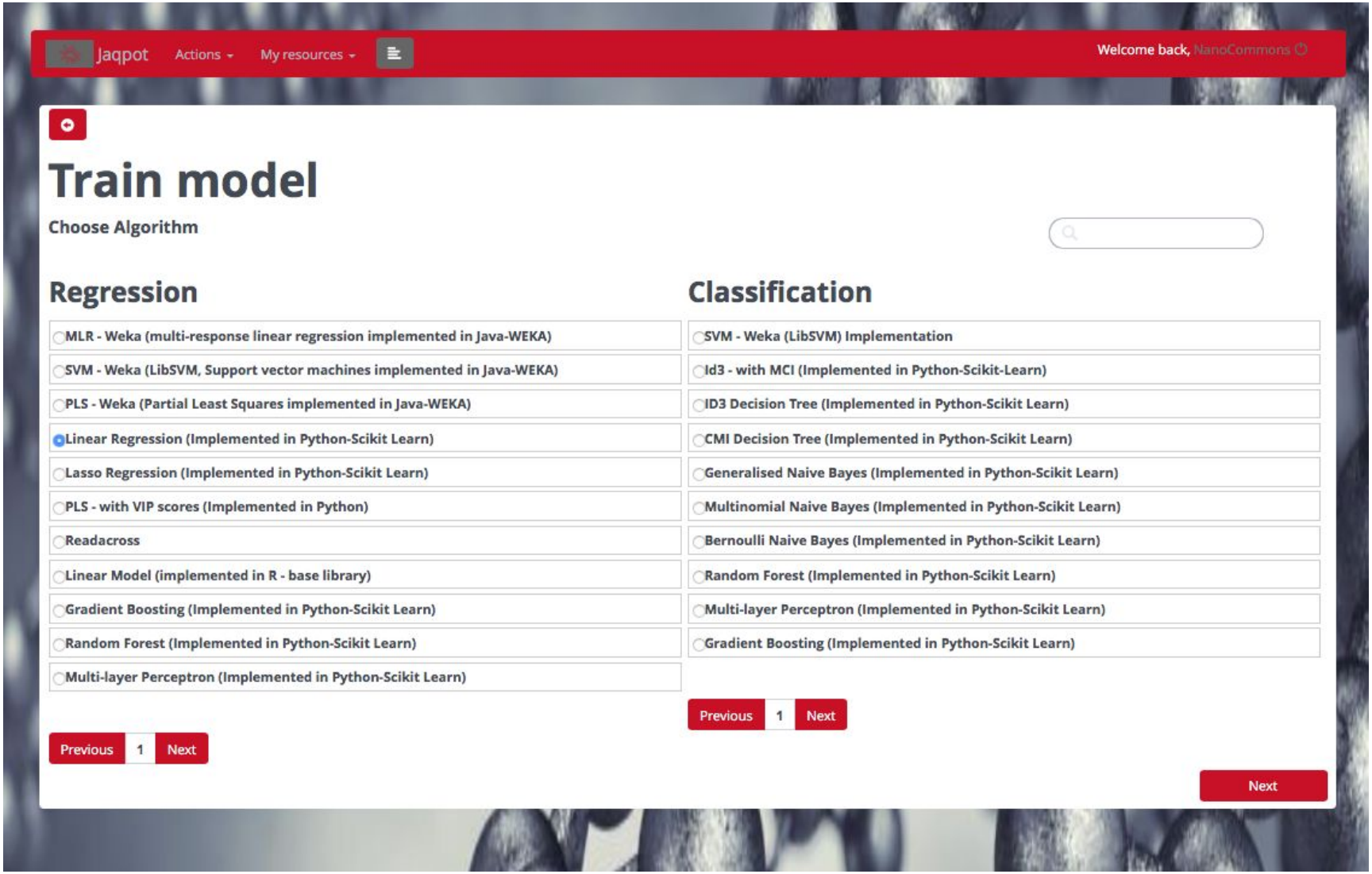
Jaqpot 4


Jaqpot
Sign in
Create Account
Create Dataset
Select substances, properties and descriptors.
Create Dataset
NanoQSAR validation schemes
Validate model using different validation options.
Cross
Data Split
External
R² = 0.87
Y
X
NanoQSAR modelling
Train a model or make a prediction.
Train
Predict
Optimal Experimental Design
Optimise the design of the experiments.
Iterative Experimental Design
Factorial Design
Interlaboratory Comparison
Perform statistical laboratory quality control.
Interlaboratory Proficiency Testing
Read Across
Make a prediction based on the read across method on an existing dataset.
Train
Predict
SOURCE
DOCUMENTATION
ISSUES
Copyright © 2015 Jaqpot.All rights reserved.
Unit of Process Control and Informatics
School of Chemical Engineering
National Technical University of Athens

My resources



Model training - Choice of algorithm

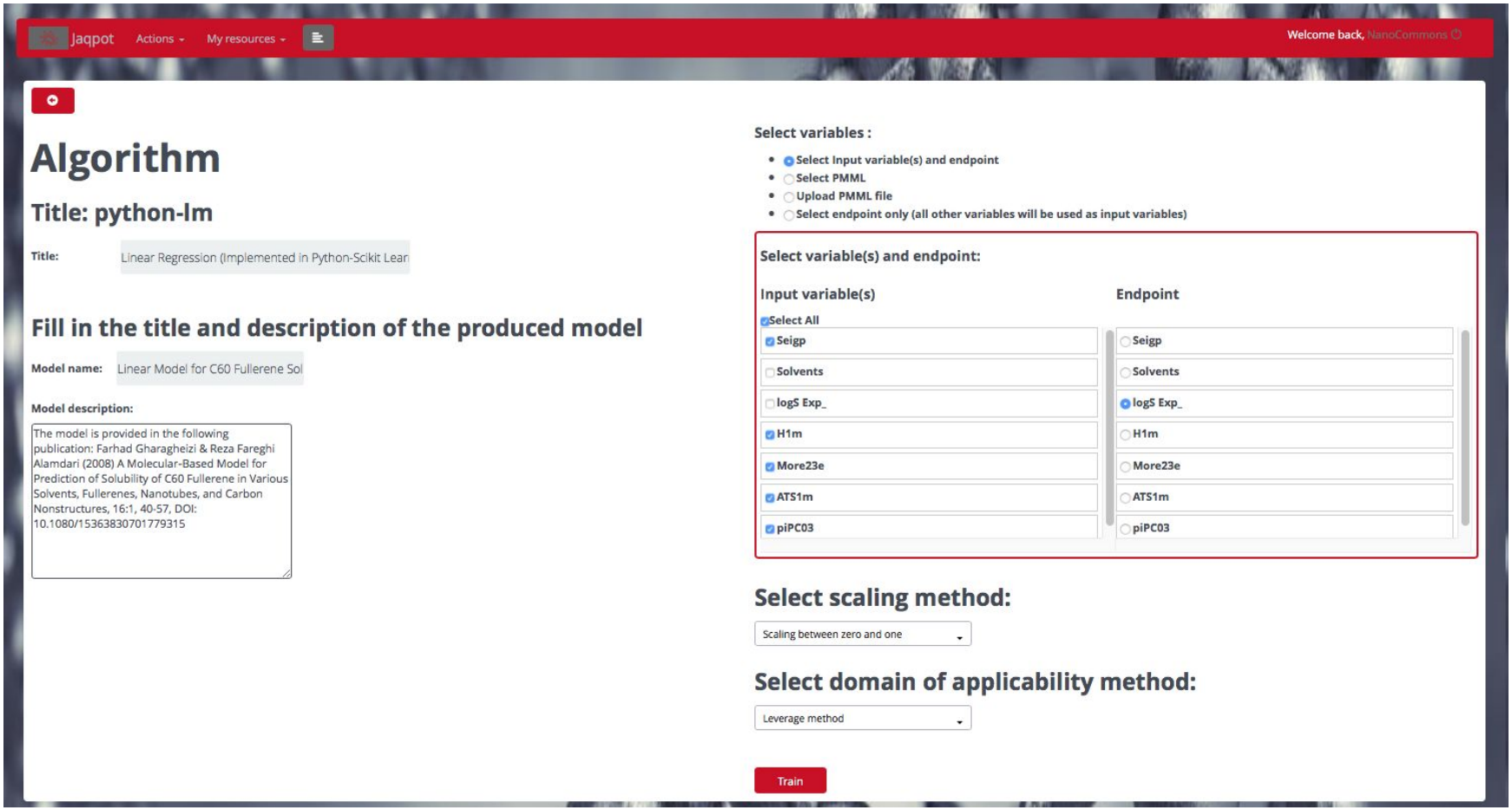


Ontological annotation of algorithms

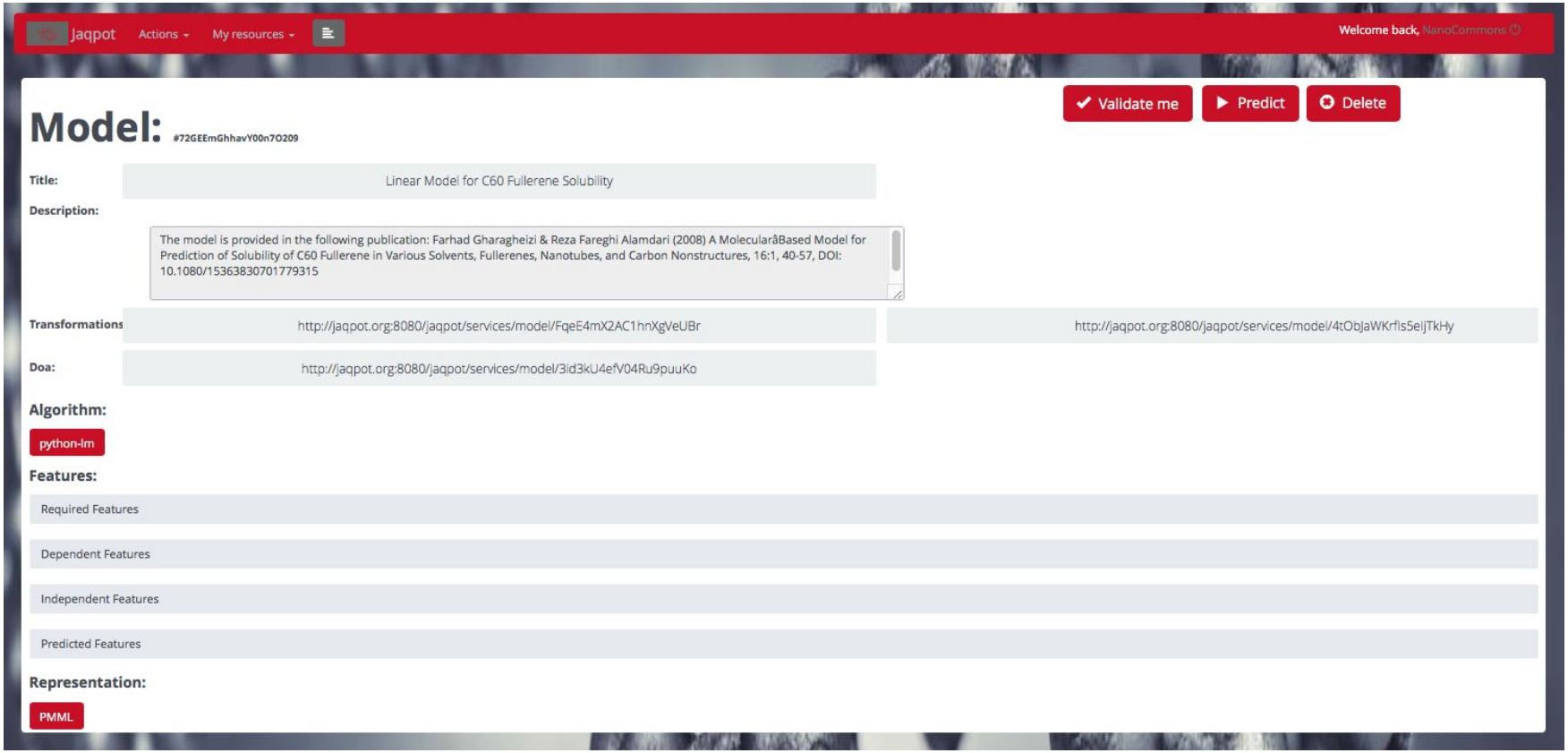
	Category / URI	Description	OpenTox Ontological Classes
	Preparation-additional		
1	http://jaqpot.org:8080/jaqpot/services/algorithm/scaling	Scaling	"ot:Algorithm", "ot:Scaling", "ot:Transformation"
	WEKA (Java)		
1	http://jaqpot.org:8080/jaqpot/services/algorithm/weka-mlr	MLR - Weka (multi-response linear regression implemented in Java-WEKA)	"ot:Algorithm", "ot:Regression", "ot:SupervisedLearning"
	Python		
1	http://jaqpot.org:8080/jaqpot/services/algorithm/python-id3-mci	Id3 - with MCI (Implemented in Python-Scikit-Learn)	"ot:Algorithm", "ot:Classification", "ot:SupervisedLearning"

OpenTox Algorithm Ontology http://old.opentox.org/data/documents/development/RDF%20files/AlgorithmTypes

Model training - Algorithm parameters

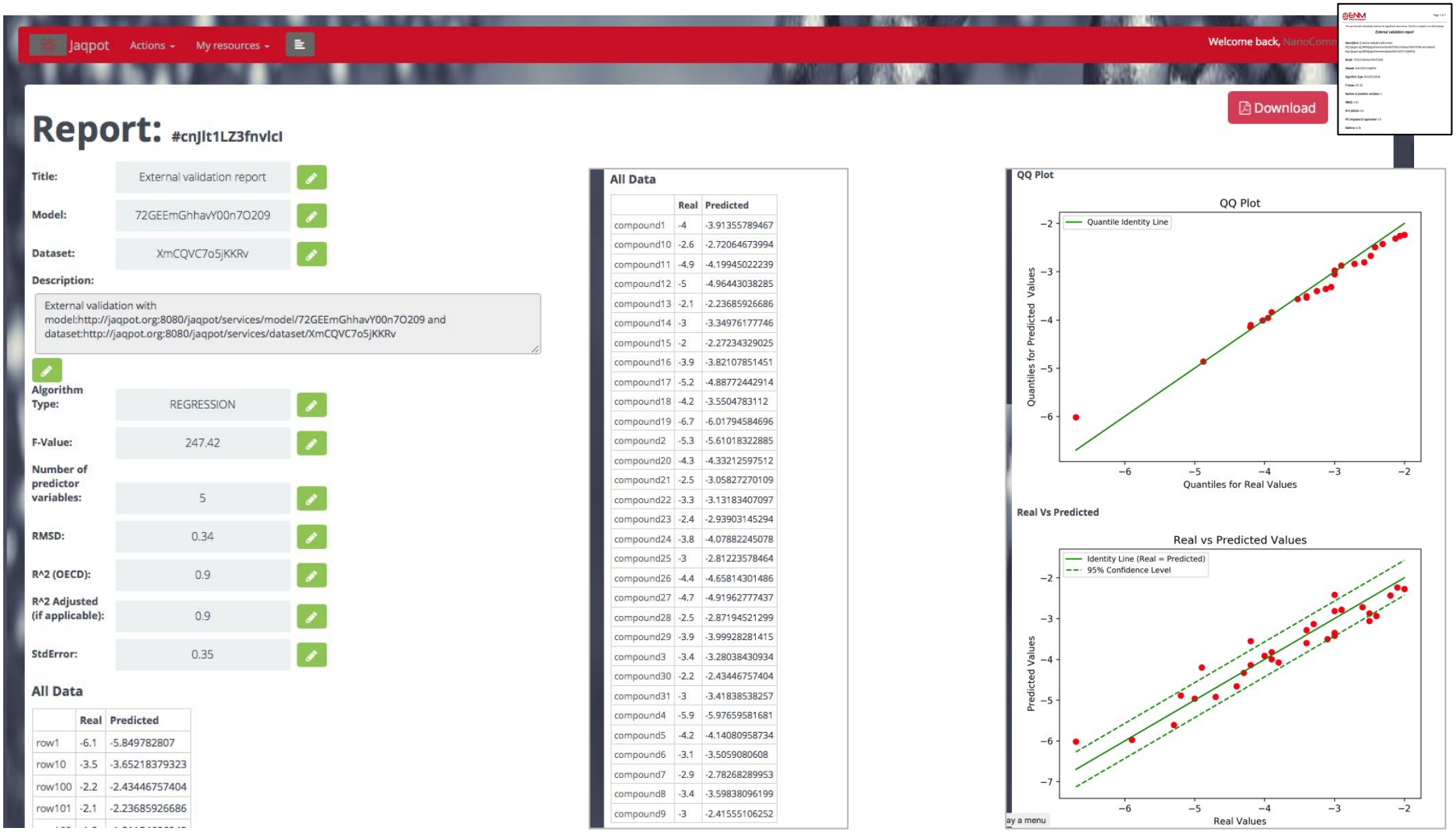


Model webpage



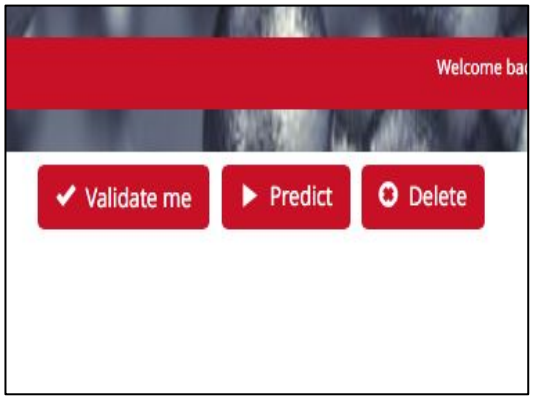
URI of model: http://www.jaqpot.org/m_detail?name=72GEEmGhhavY00n7O209

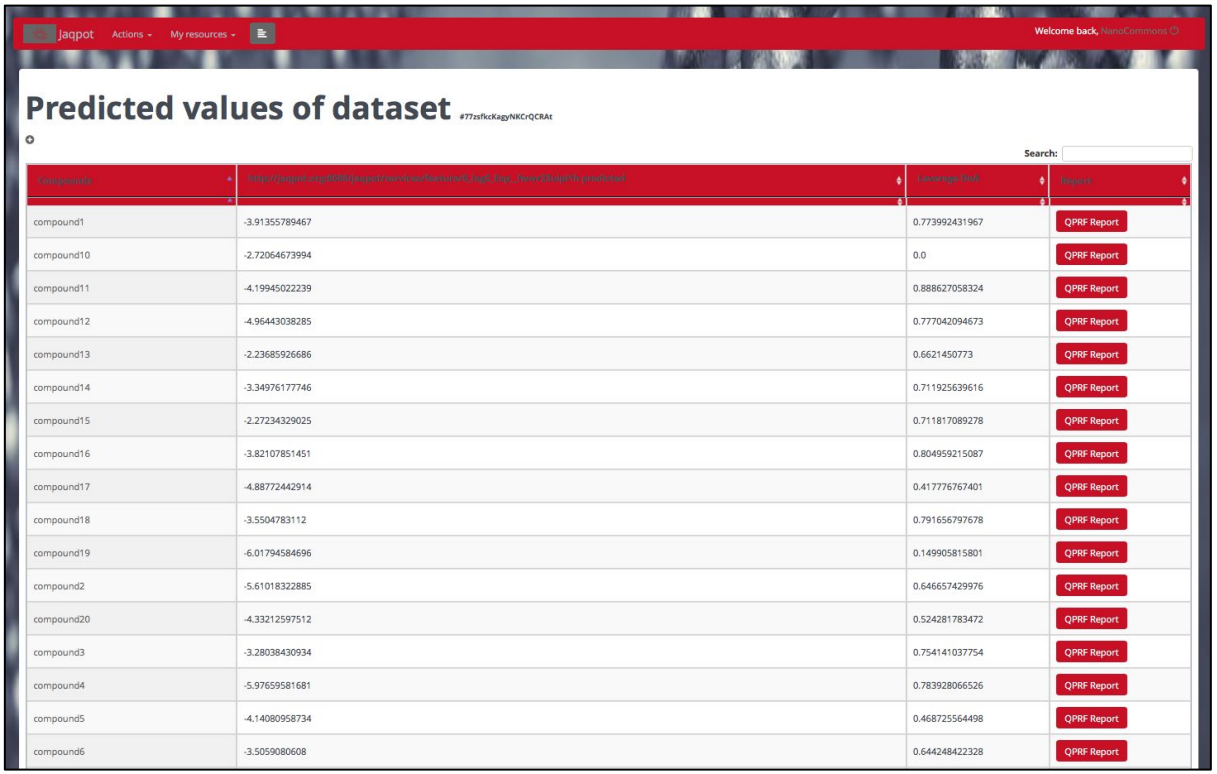
Validation report



http://www.jaqpot.org/report?name=cnJlt1LZ3fnvlcl

Predictions



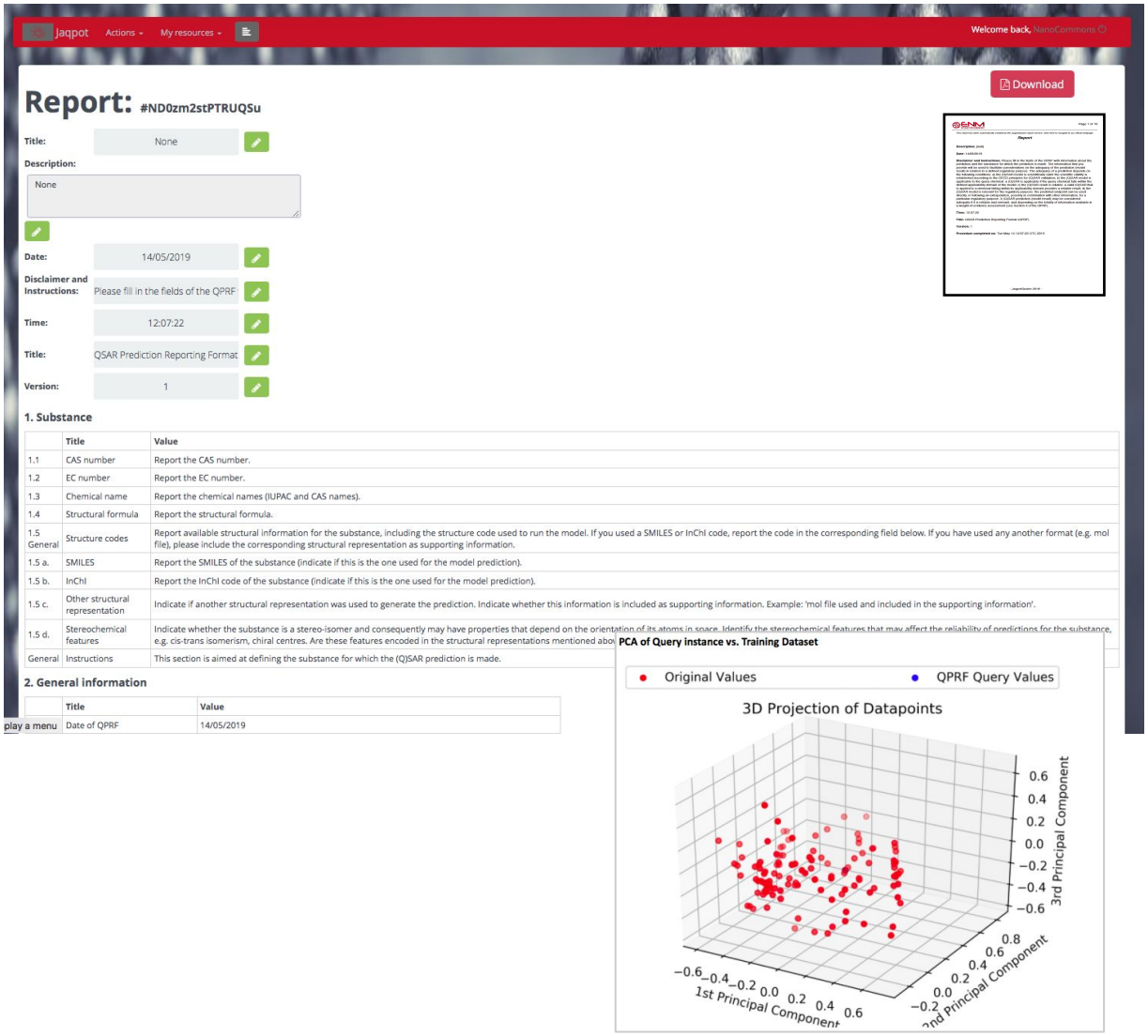


compounds		Leverage MIA	Report
compound1	-3.91355789467	0.773992431967	QPRF Report
compound10	-2.72064673994	0.0	QPRF Report
compound11	-4.19945022239	0.888627058324	QPRF Report
compound12	-4.96443038285	0.777042094673	QPRF Report
compound13	-2.23685926686	0.6621450773	QPRF Report
compound14	-3.34976177746	0.711925639616	QPRF Report
compound15	-2.27234329025	0.711817089278	QPRF Report
compound16	-3.82107851451	0.804959215087	QPRF Report
compound17	-4.88772442914	0.417776767401	QPRF Report
compound18	-3.5504783112	0.791656797678	QPRF Report
compound19	-6.01794584696	0.149905815801	QPRF Report
compound2	-5.61018322885	0.646657429976	QPRF Report
compound20	-4.33212597512	0.524281783472	QPRF Report
compound3	-3.28038430934	0.754141037754	QPRF Report
compound4	-5.97659581681	0.783928066526	QPRF Report
compound5	-4.14080958734	0.468725564498	QPRF Report
compound6	-3.5059080608	0.644248422328	QPRF Report

QPRF report

The QPRF (QSAR prediction reporting format) report generated by Jaqpot contains all the fields required by the OECD guidelines, namely:

- Substance
  - Contains information such as CAS and EC numbers, SMILES, InChi, etc.
- General
  - Information such as date and creator name and email.
- Prediction
  - Biological endpoint, variables, model used, DoA, etc
- Adequacy
  - Optional field, containing regulatory purpose, conclusion etc



Read Across

- Users select **distance metric** and a **threshold** between 0 and 1 (the closest to 0 the strictest) used to assess nanoparticle similarity.
- Optionally select a method to calculate the **prediction confidence level** (Currently “Nearest Neighbour Confidence”)

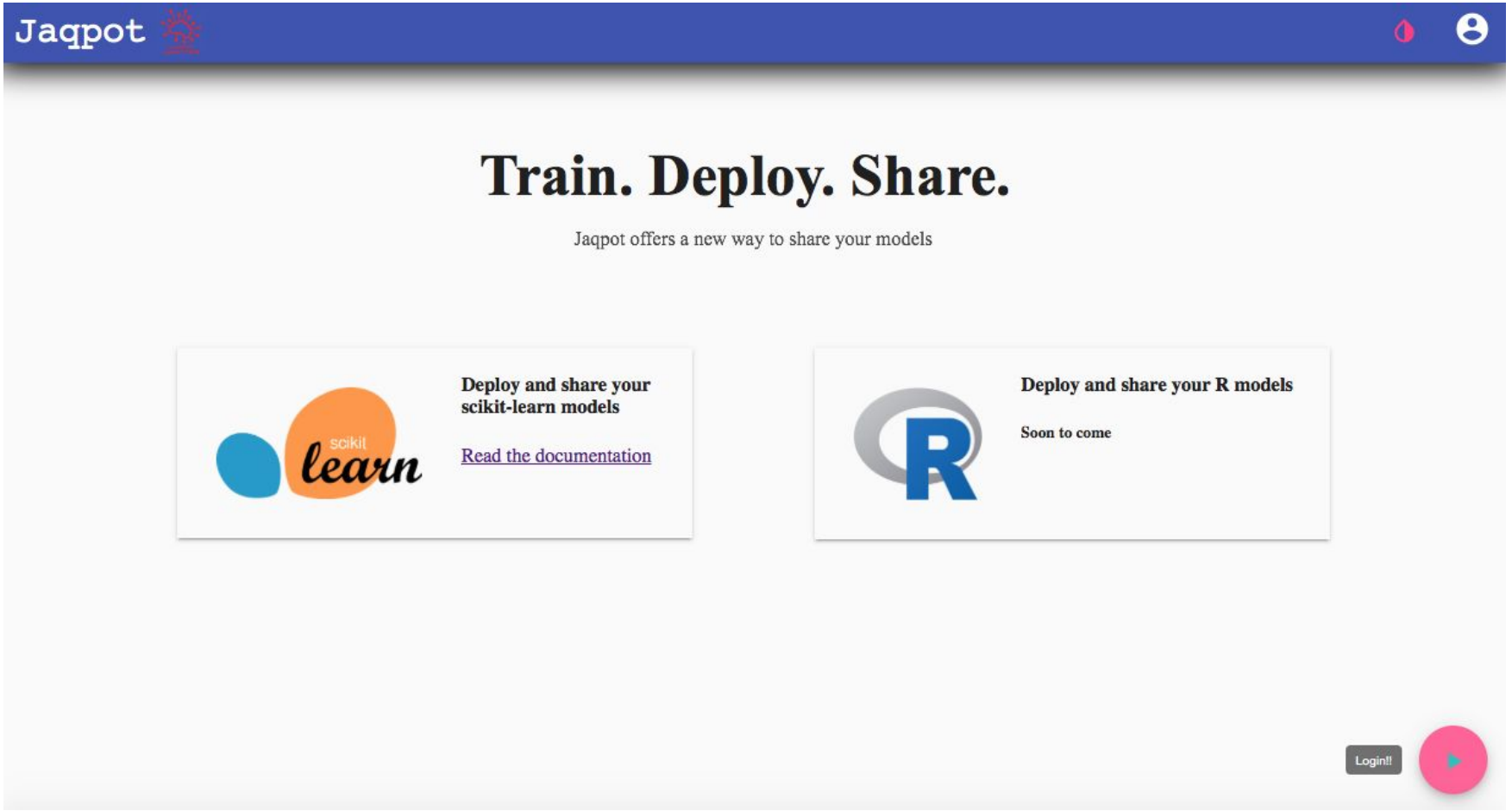
Predicted values of dataset #ghh9UagIv2ZuOjJaaZHA

Search:

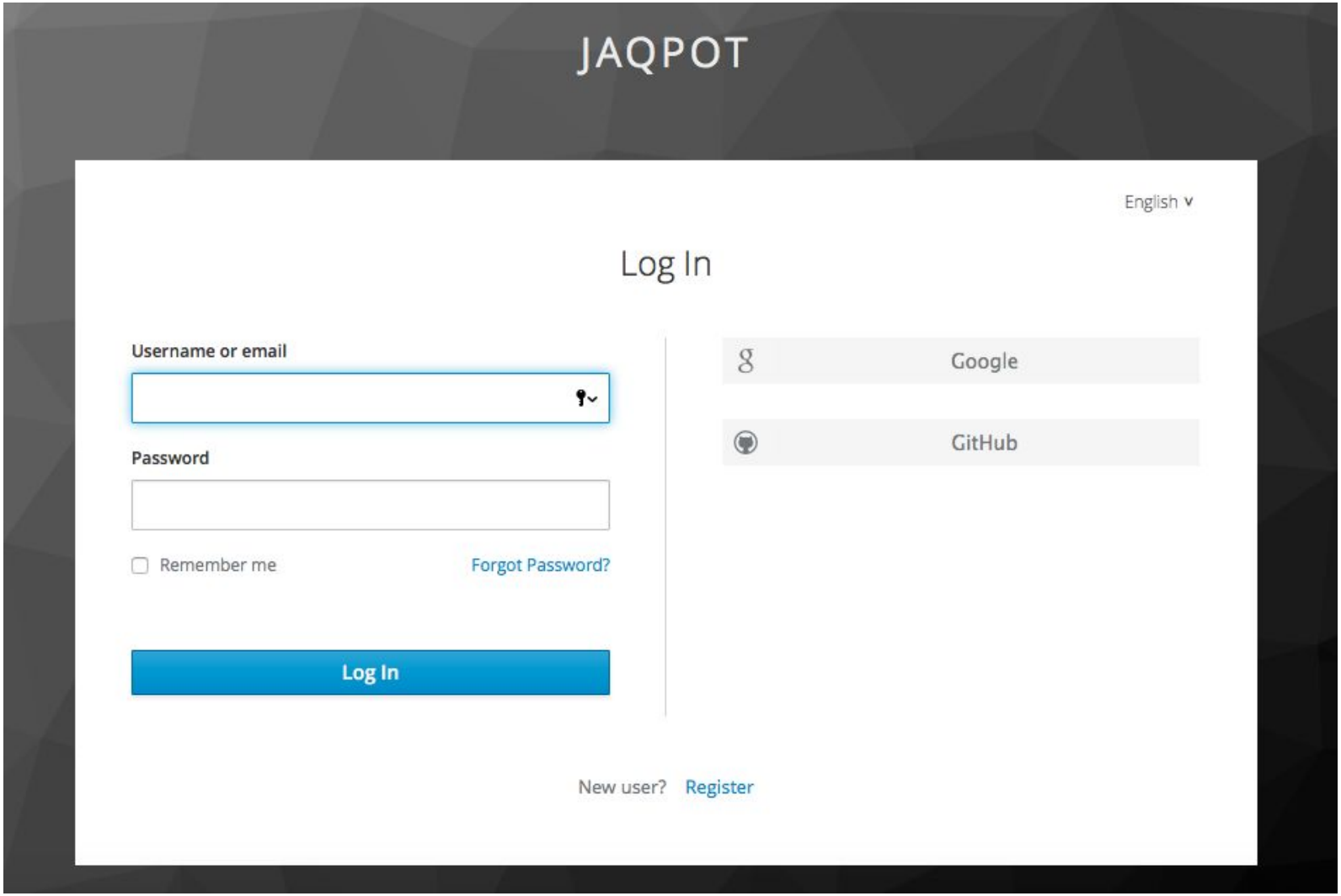
Compounds	https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed/94D664CFE4928A0F400A5AD8CA753B52E049A688/3ed642f0-1b42-387a-9966-dea5b91e5f8a manhattan	https://apps.ideaconsult.net/enmtest/property/TOX/UNK 1b42-387a-9966-dea5b91e5f8a confidence
G15.AC_1	-5.42	0.93
G15.AHT_1	-1.01	1.0
G15.Asn-SH_1	-5.49	0.93
G15.AUT_1	-1.32	1.0
G15.CIT_1	-5.64	0.93
G15.CTAB_1	-5.86	1.0
G15.DDT@CTAB_1	-4.75	0.95

Jaqpot 5

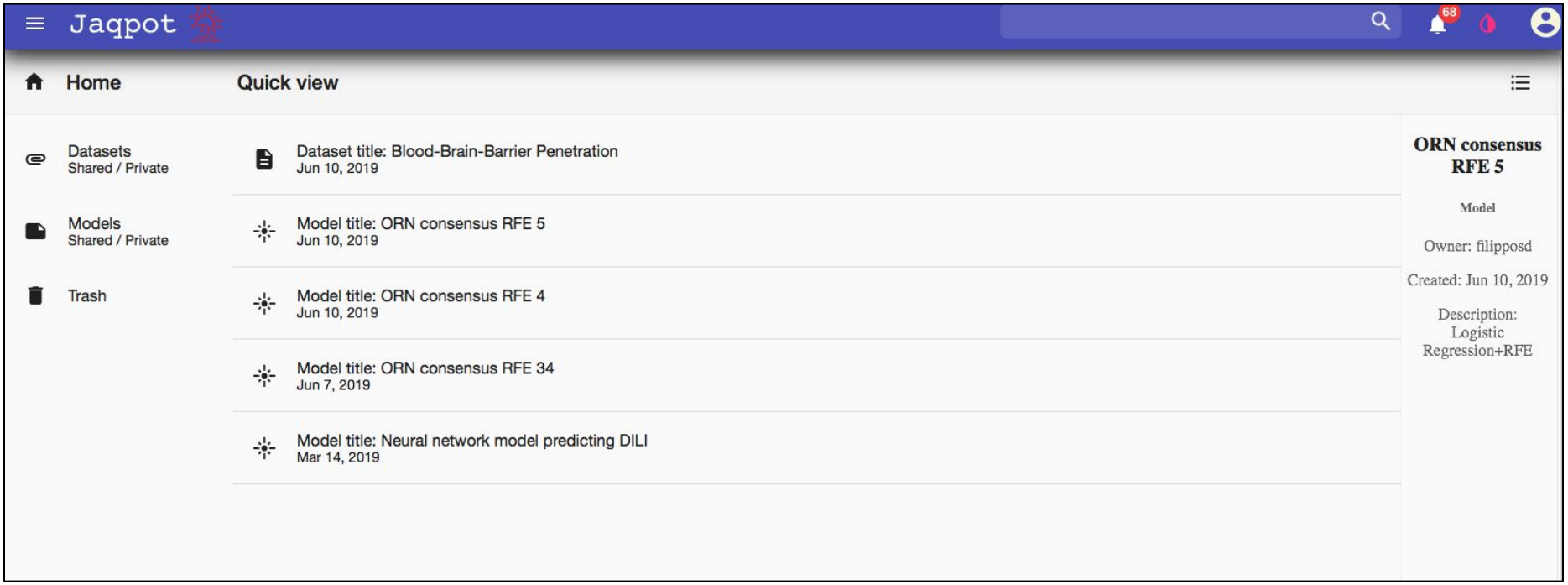
Jaqpot 5


Jaqpot
Train. Deploy. Share.
Jaqpot offers a new way to share your models
Deploy and share your scikit-learn models
Read the documentation
Deploy and share your R models
Soon to come
Login!!

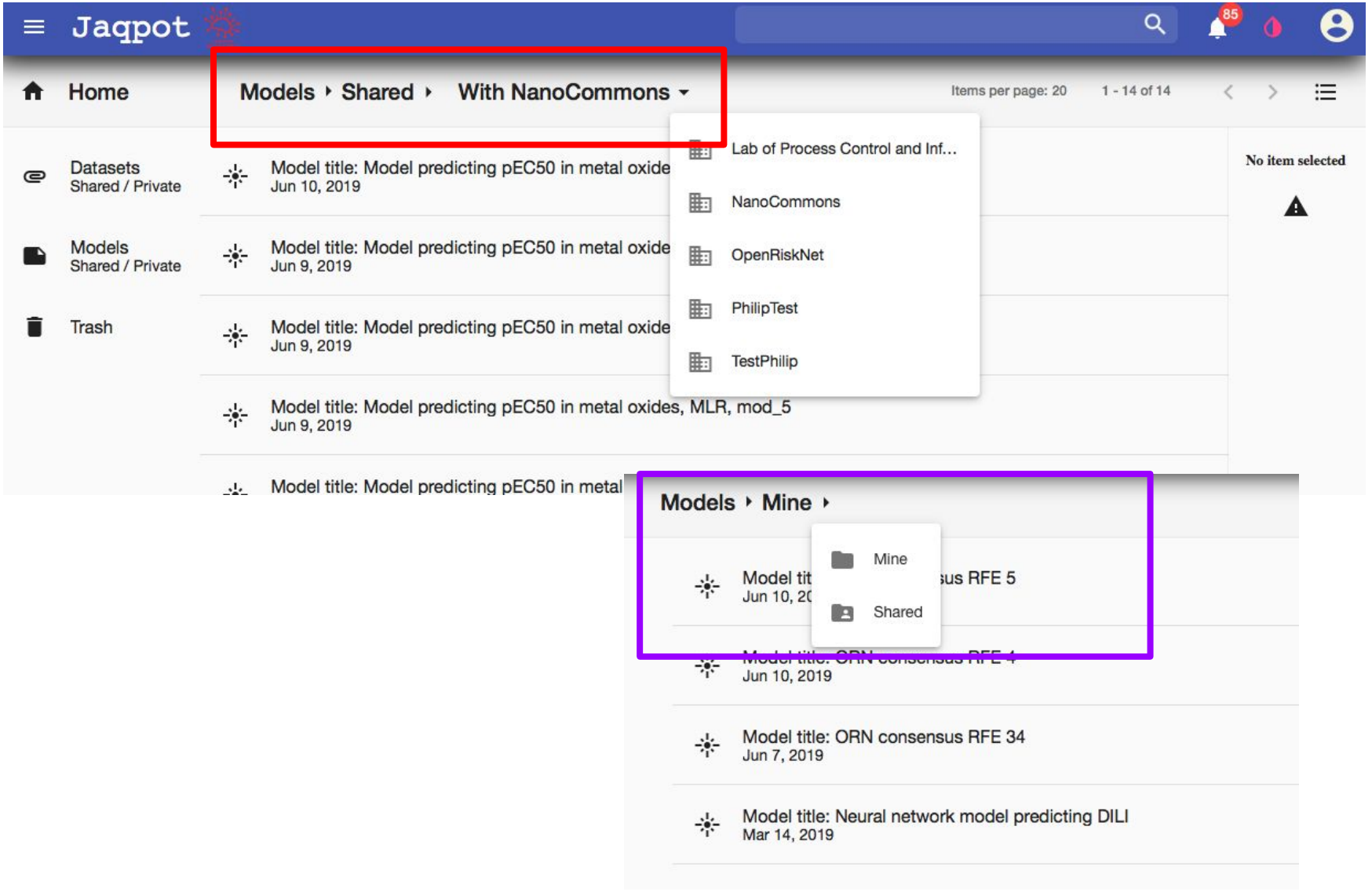
Jaqpot 5 - Login



Jaqpot 5 - Home



Jaqpot 5 - Shared space - Models



https://app.jaqpot.org/model/uxRBCMsV9lkSQT1Kw7km

Jaqpot 5 - Model page

Jaqpot

Overview | Data | Predict / Validate | Discussion

MODEL

Title: Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Owner: hsarimv

Description:

Neural Model

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

The model is provided in the following publication: Farhad Gharagheizi & Reza Fareghi Alamdari (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, Fullerenes, Nanotubes, and Carbon Nonstructures, 16:1, 40-57, DOI: 10.1080/15363830701779315

Full dataset is available in this link

Training dataset is available in this link

Test dataset is available in this link

A downloadable QMRF Report is available in this link

QMRF Report

1. QSAR Identifier

1.1.QSAR identifier (title):

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

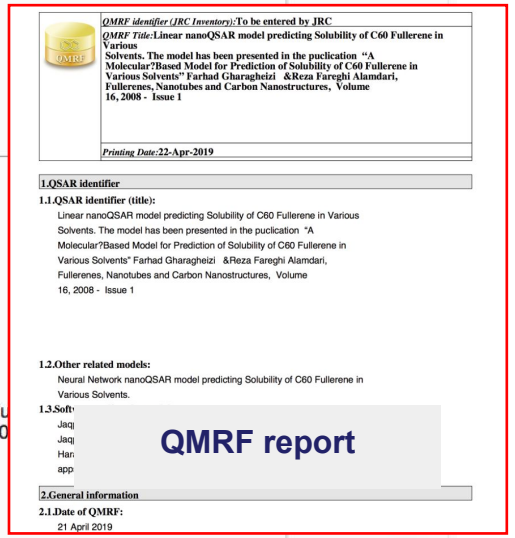
The model has been presented in the puclication "A Molecular-Based Model for Prediction of Solubility of C60 Fu
Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 200

1.2.Other related models:

Neural Network nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

1.3.Software coding the model:

Jaqpot is a web platform that support development, validation and sharing of QSAR models apps.jaqpot.org


QMRF identifier (JRC Inventory): To be entered by JRC

QMRF Title: Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model has been presented in the puclication "A Molecular?Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1

Printing Date: 22-Apr-2019

1.QSAR identifier

1.1.QSAR identifier (title):

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model has been presented in the puclication "A Molecular?Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1

1.2.Other related models:

Neural Network nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

1.3.Soft

Jaq

Jaq

Har

app

2.General information

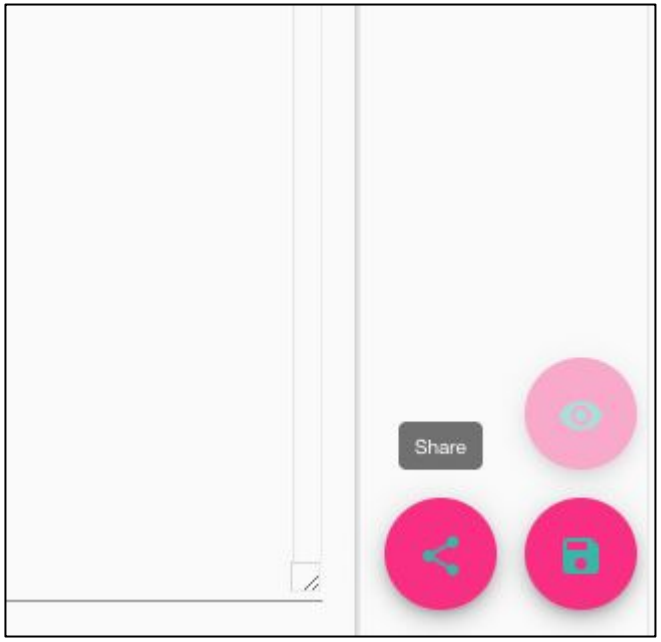
2.1.Date of QMRF:

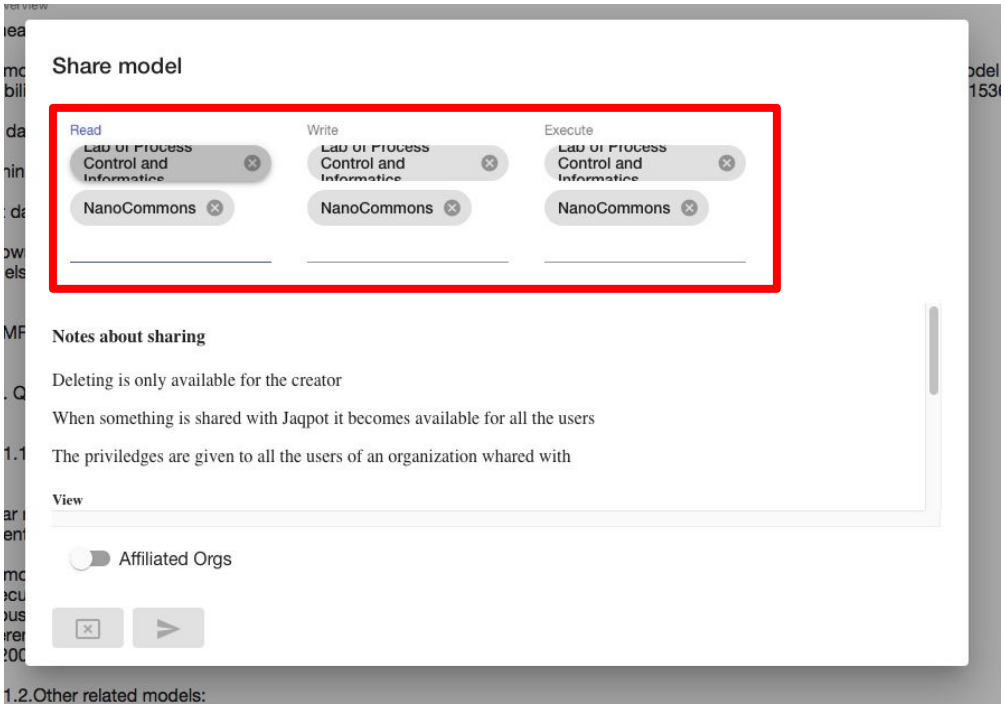
21 April 2019

QMRF report

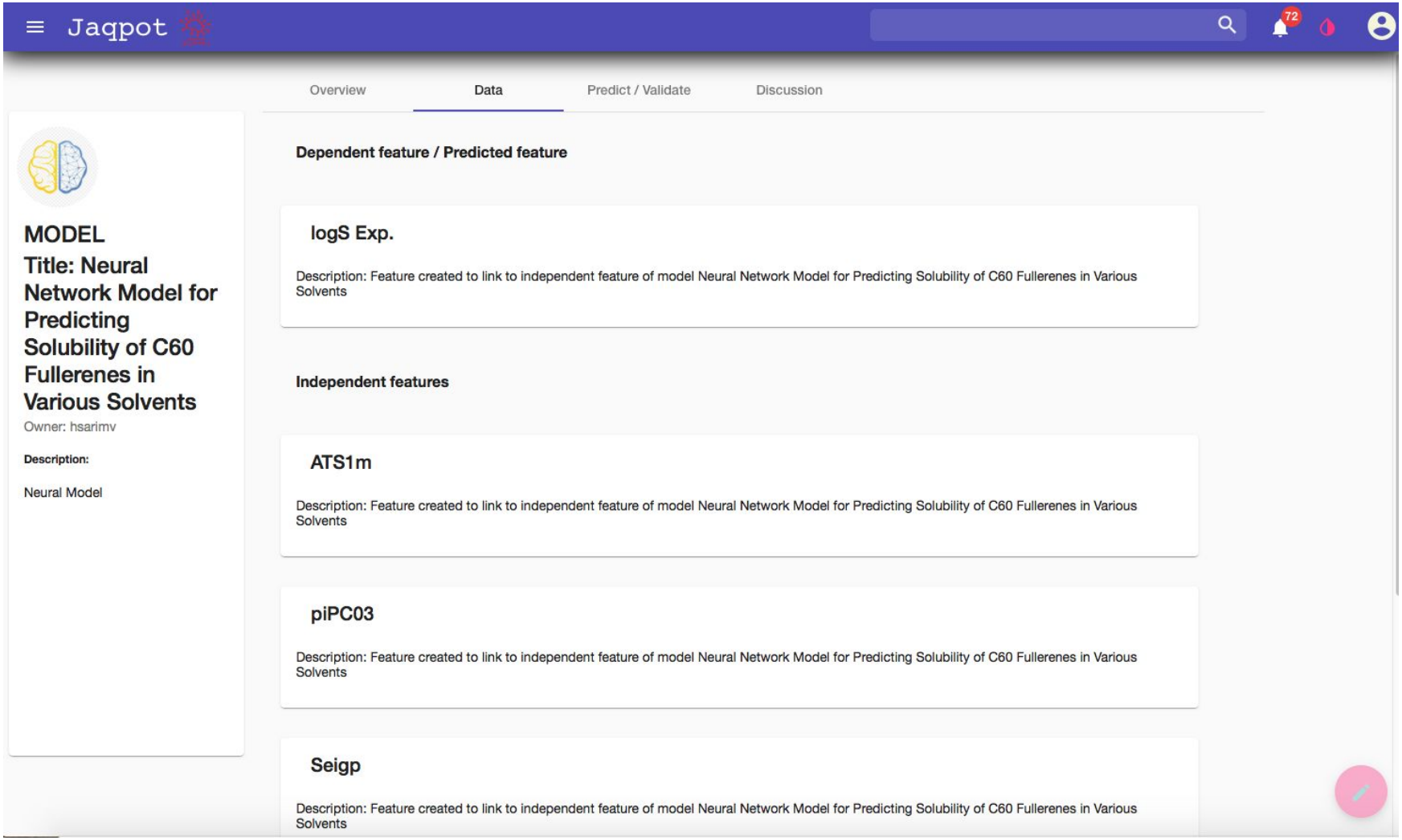
https://app.jaqpot.org/model/uxRBCMsV9lkSQT1Kw7km

Jaqpot 5 - Model page - Sharing

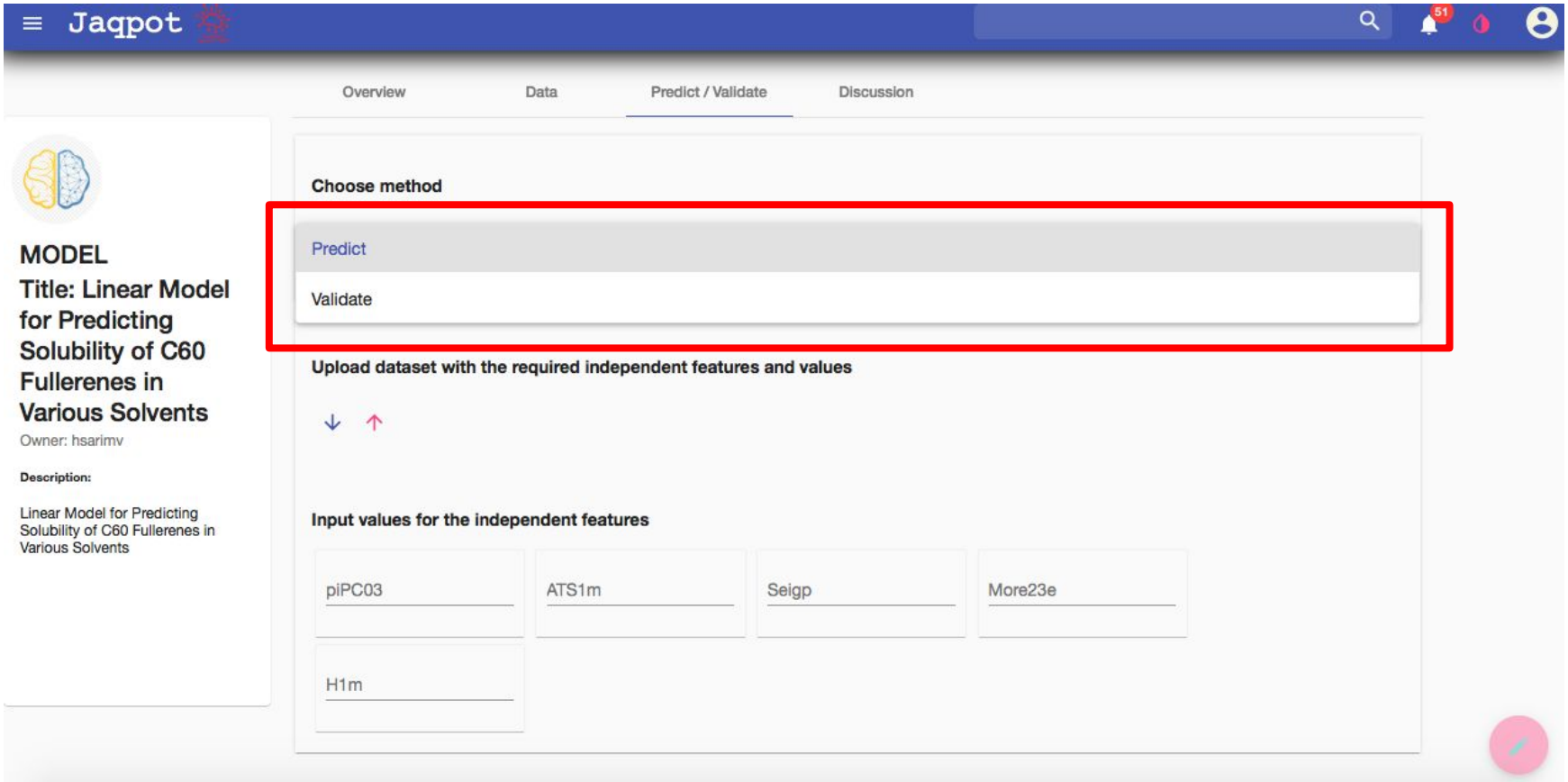




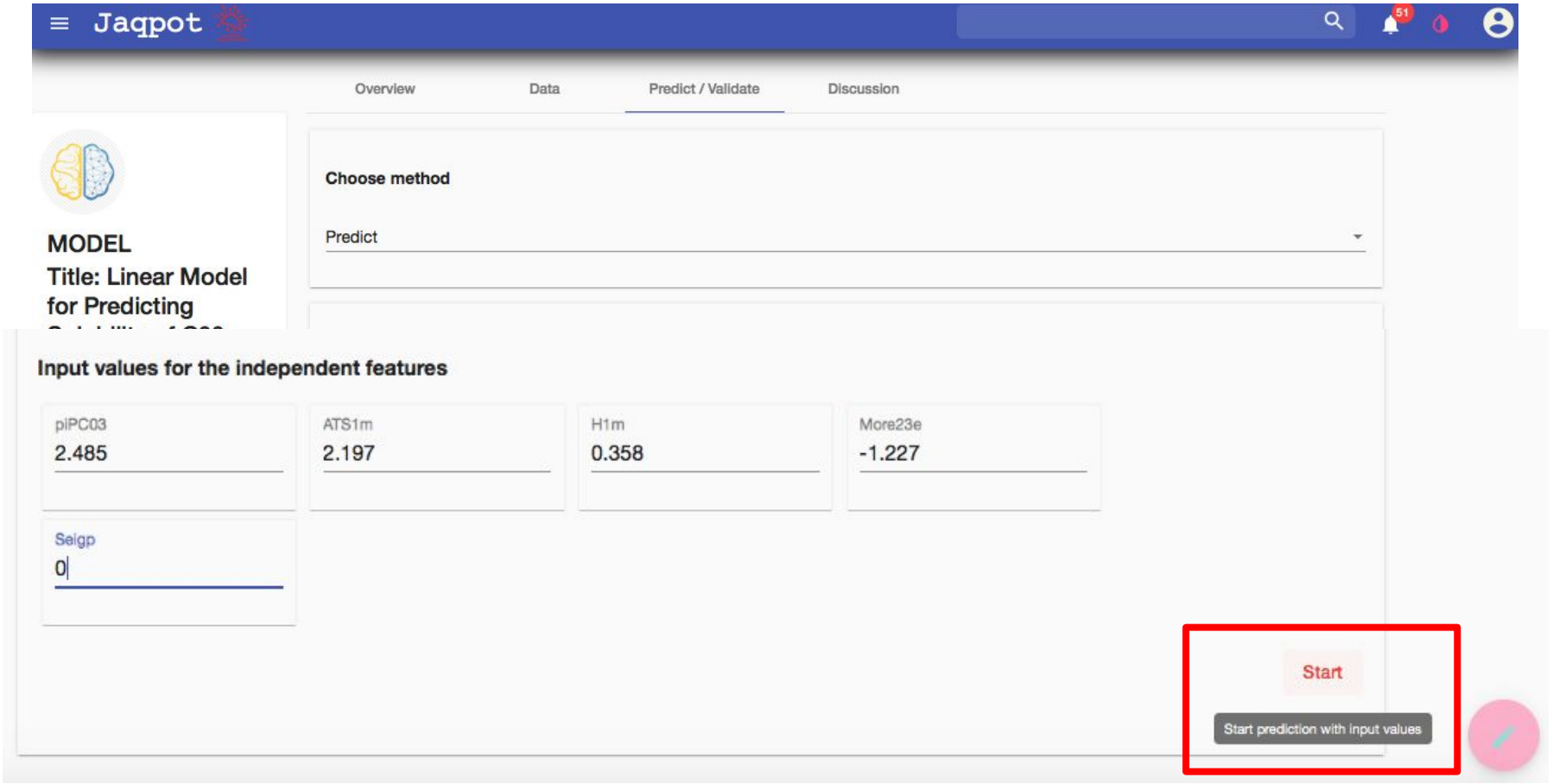
Jaqpot 5 - Model page - Data tab



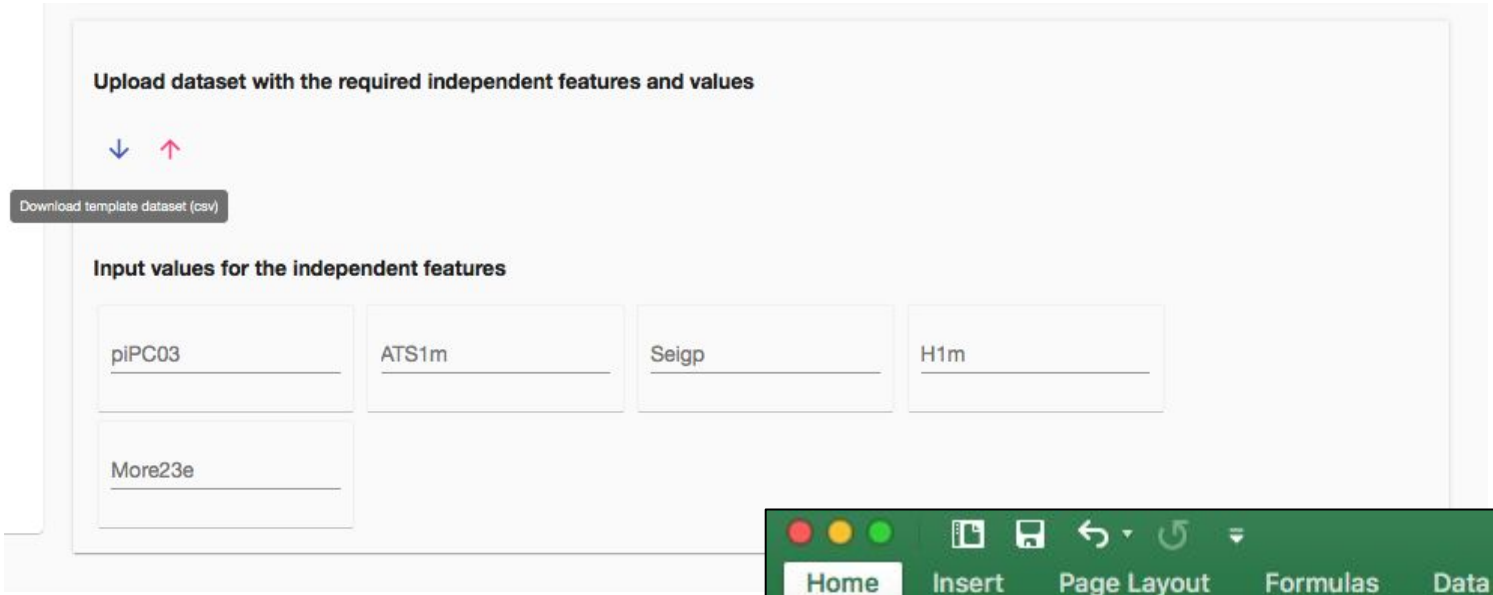
Jaqpot 5 - Model page - Predict tab



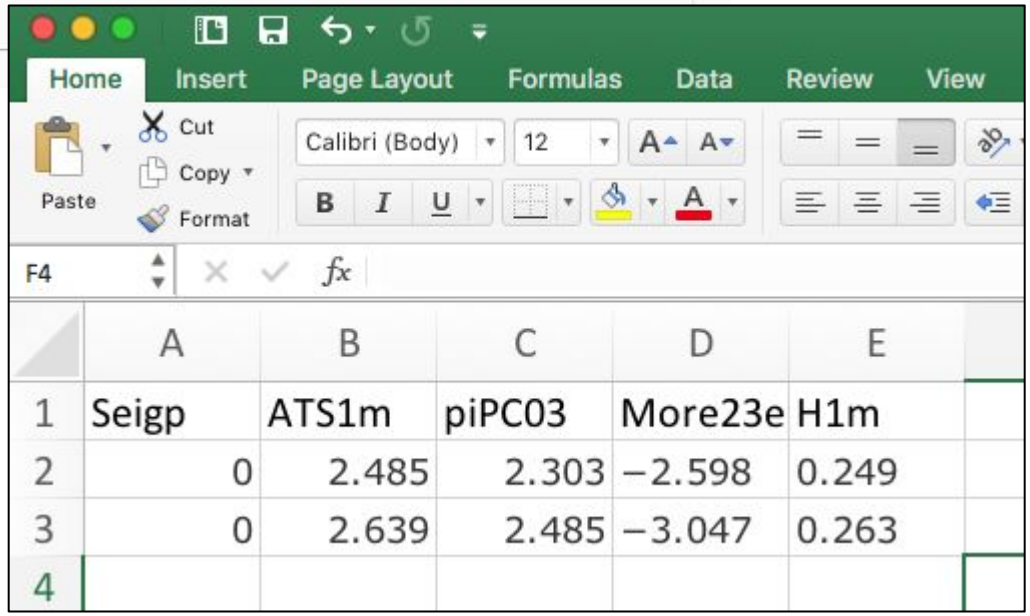
Jaqpot 5 - Predict tab - Entering values



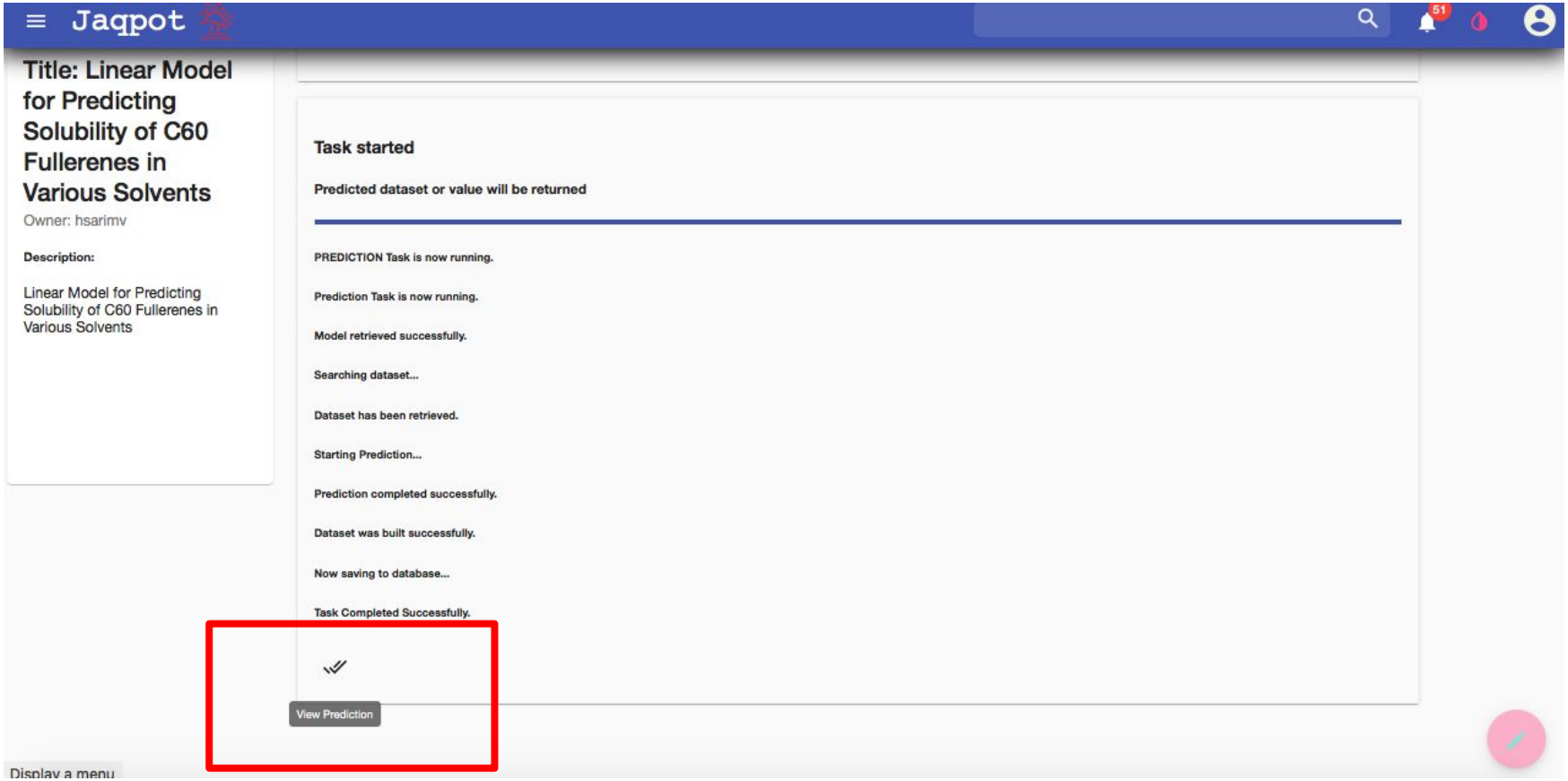
Jaqpot 5 - Predict tab – Your data via template



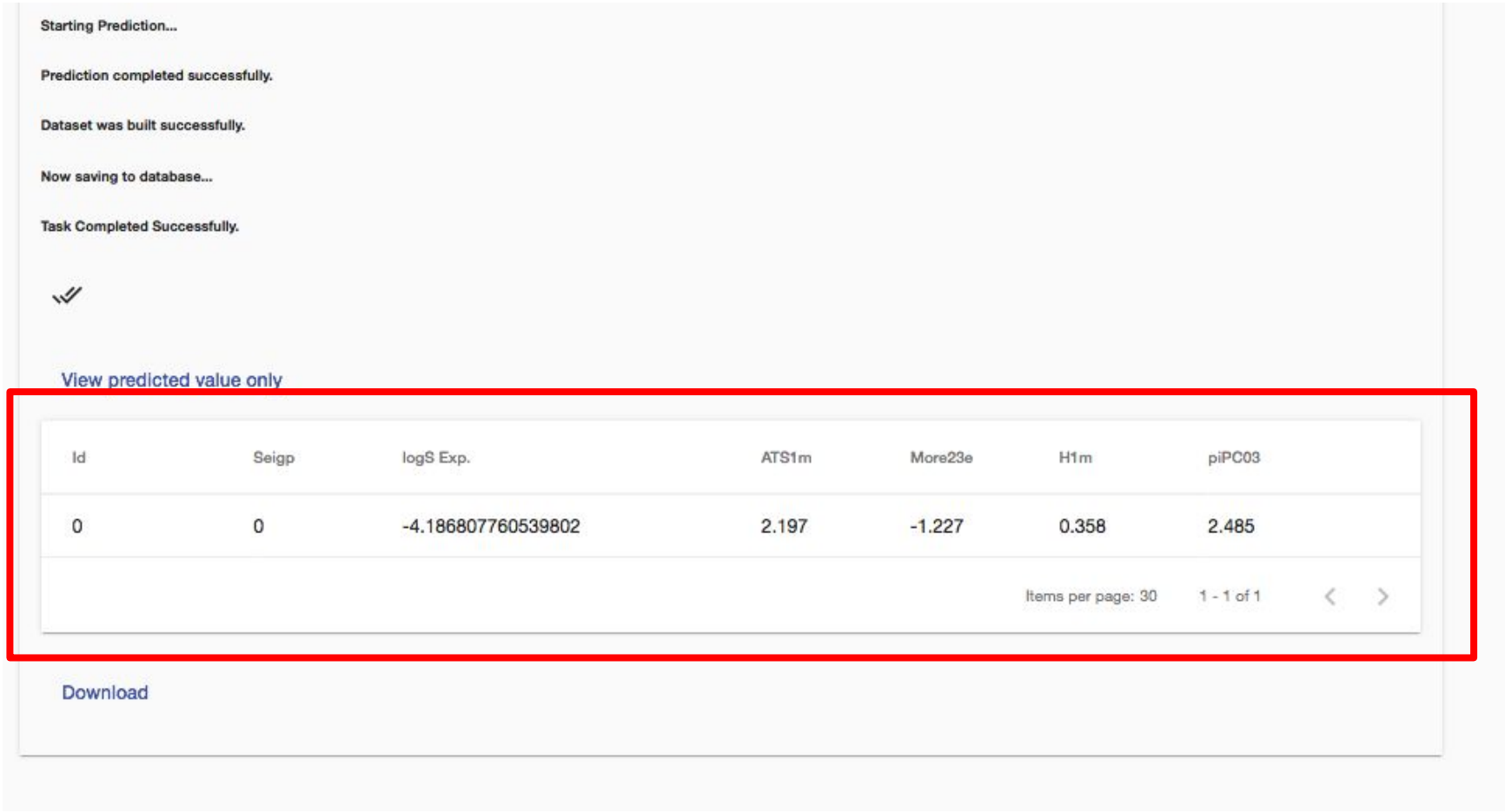
A template for users
to enter their data,
upload to Jaqpot
and get predictions.



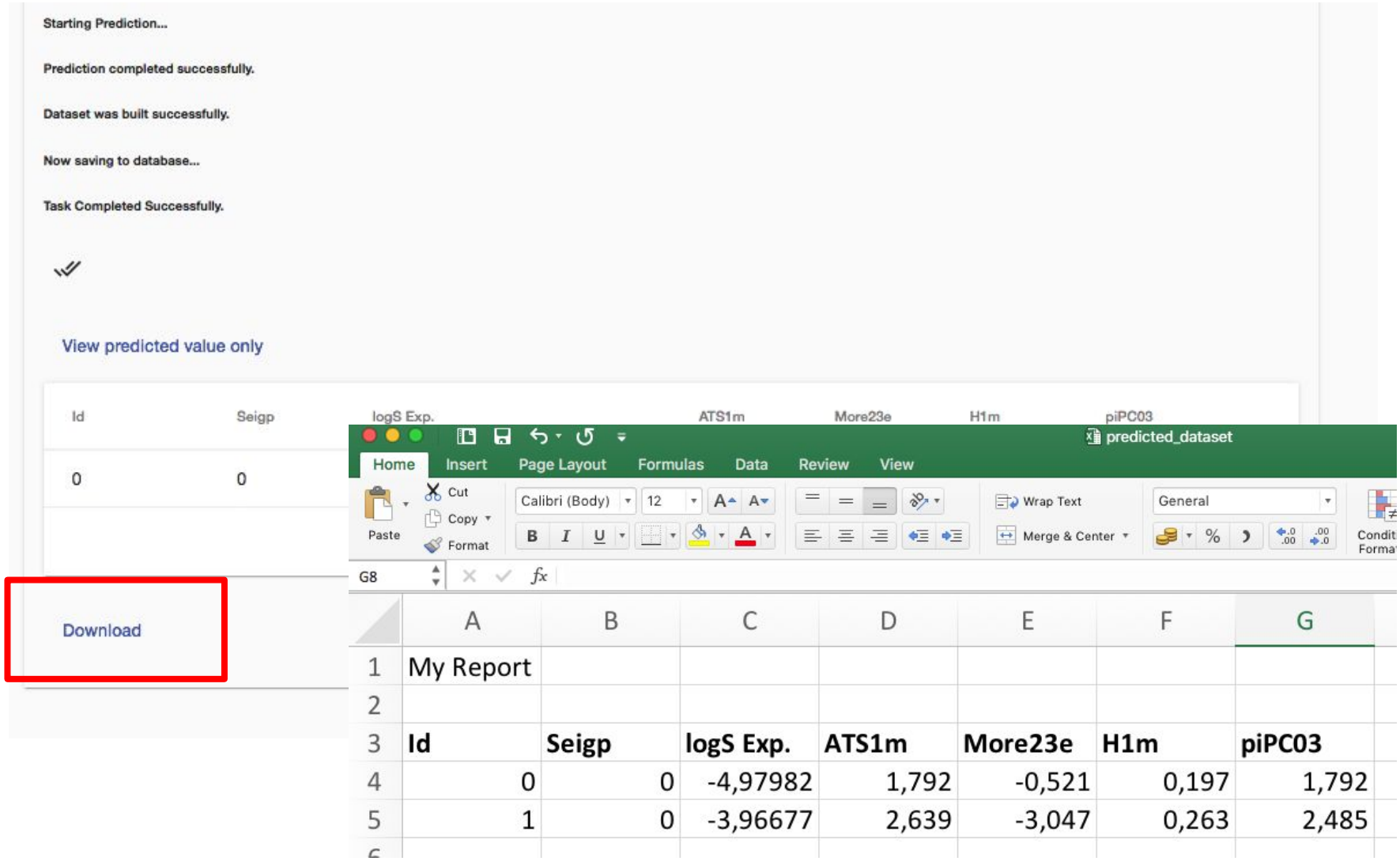
Jaqpot 5 - Model page - Predict tab



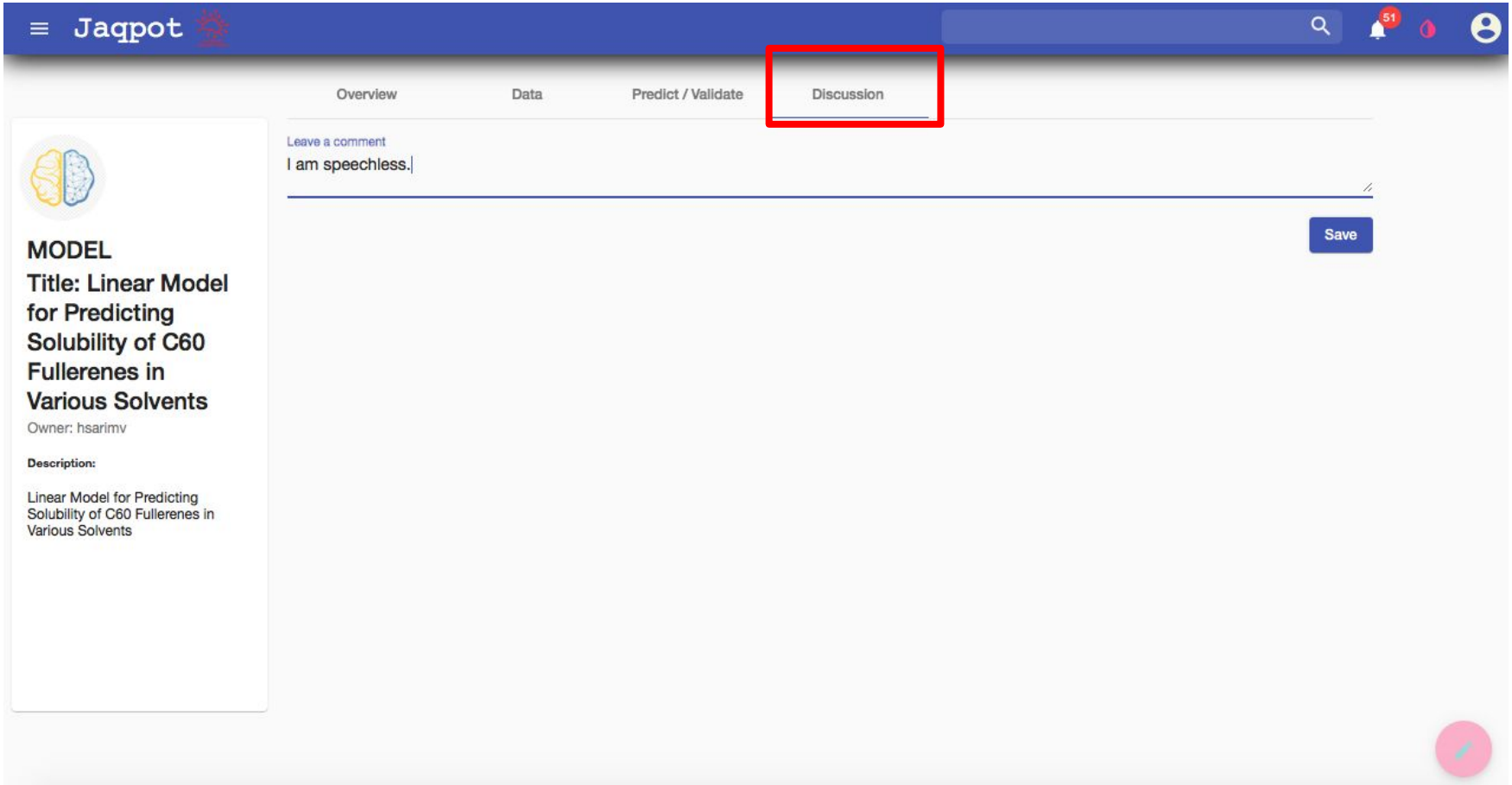
Jaqpot 5 - Predictions



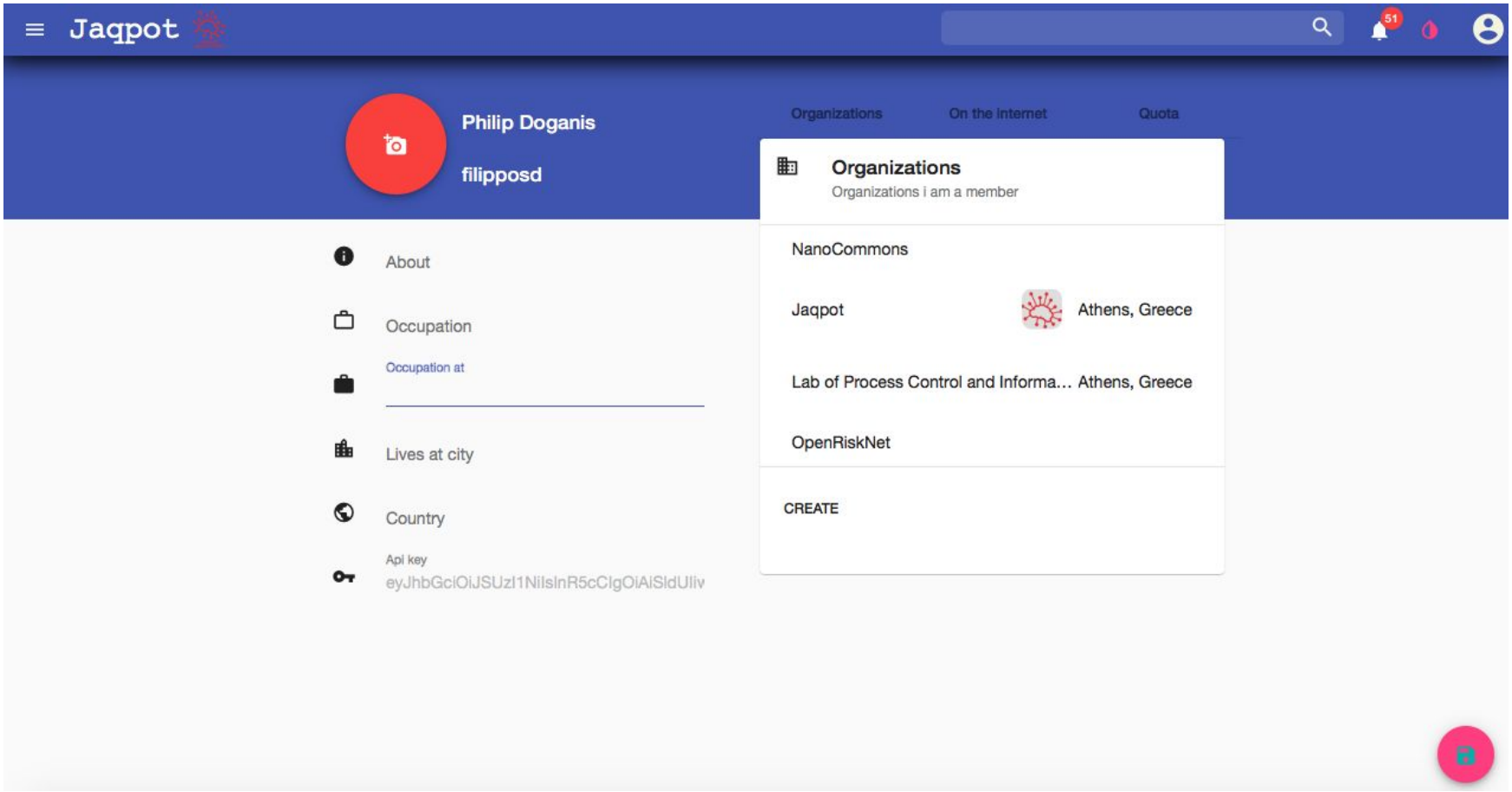
Jaqpot 5 - Predictions as CSV



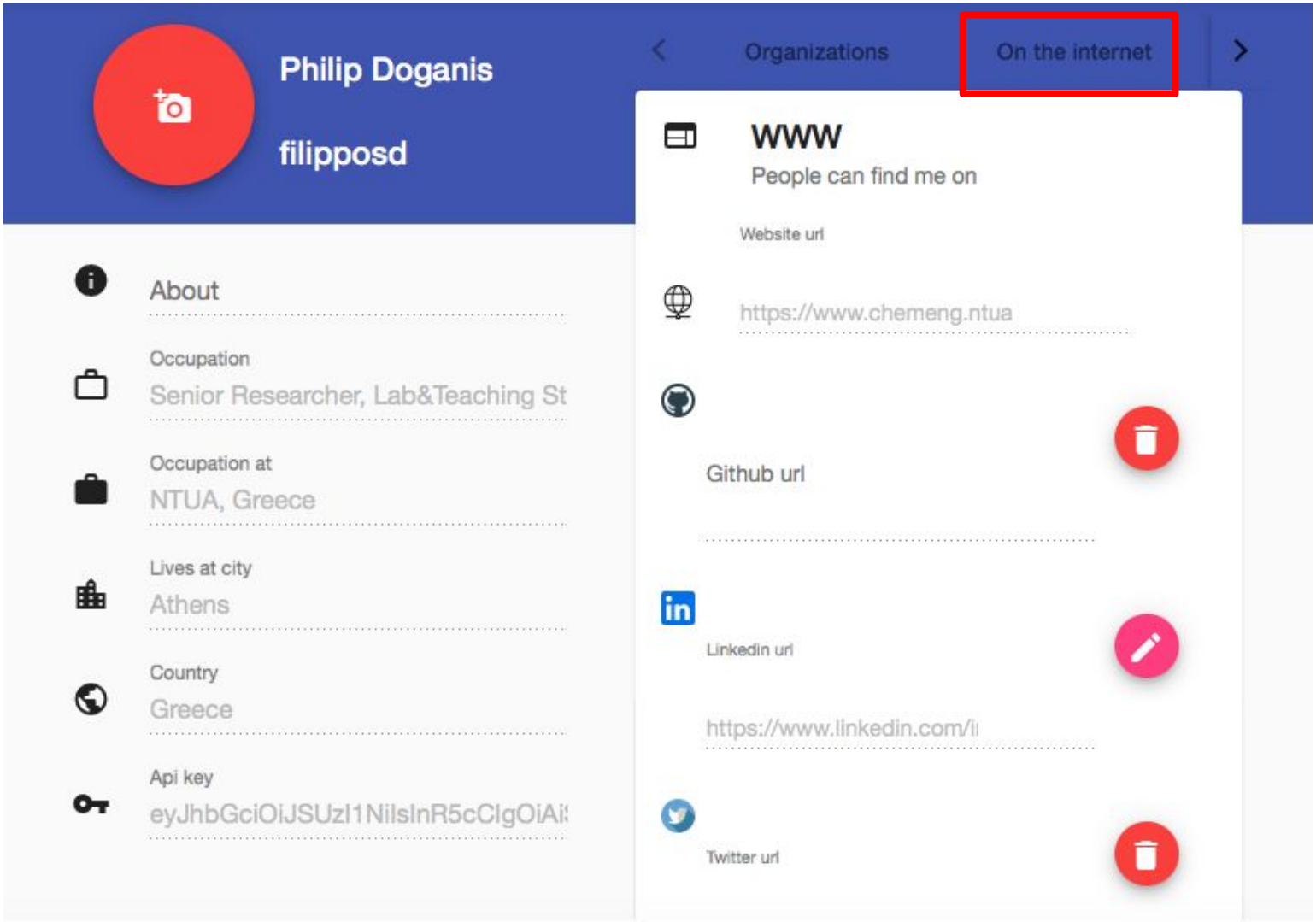
Jaqpot 5 - Social network of models



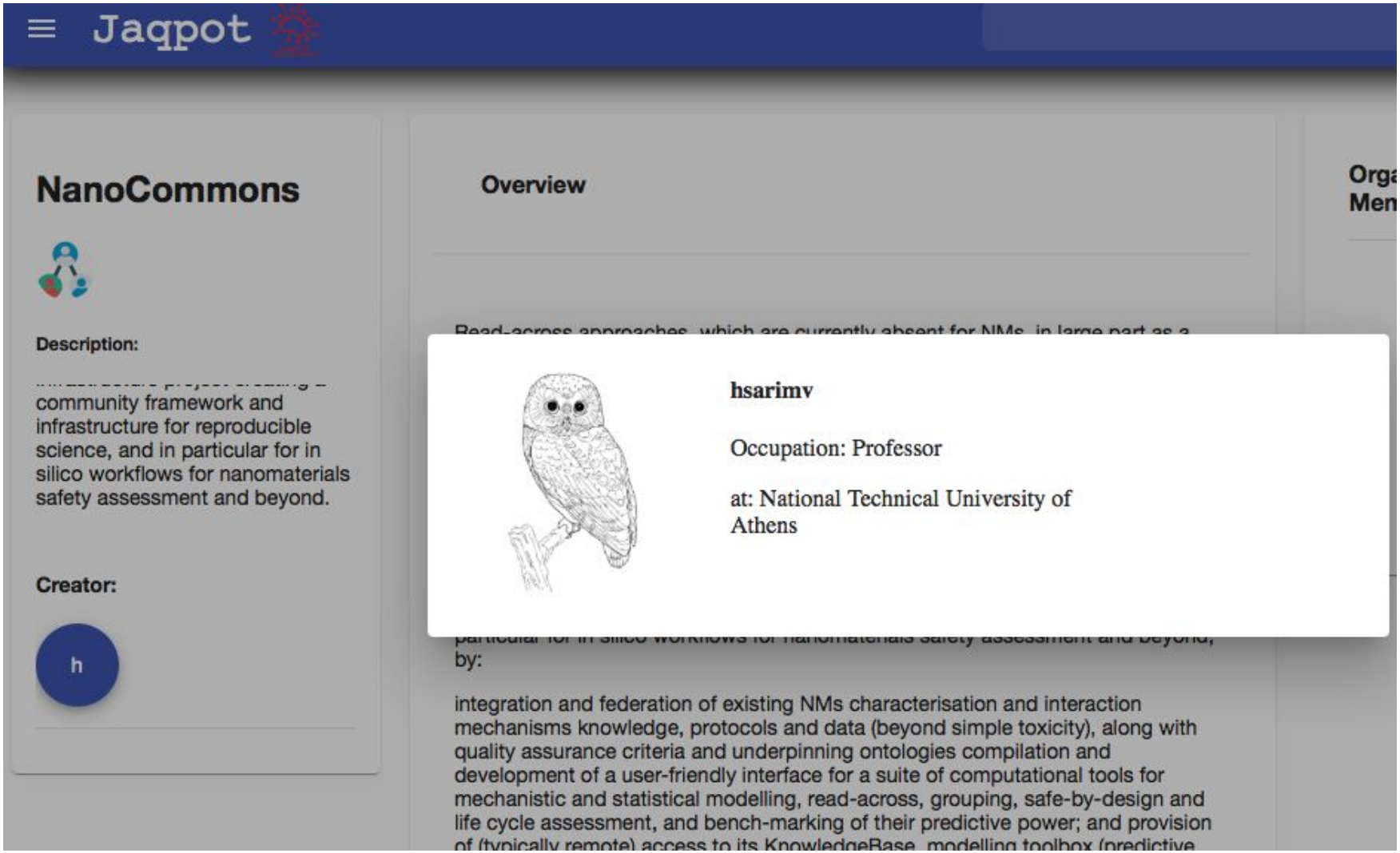
Jaqpot 5 - User profile



Jaqpot 5 - User profile - Contact Info

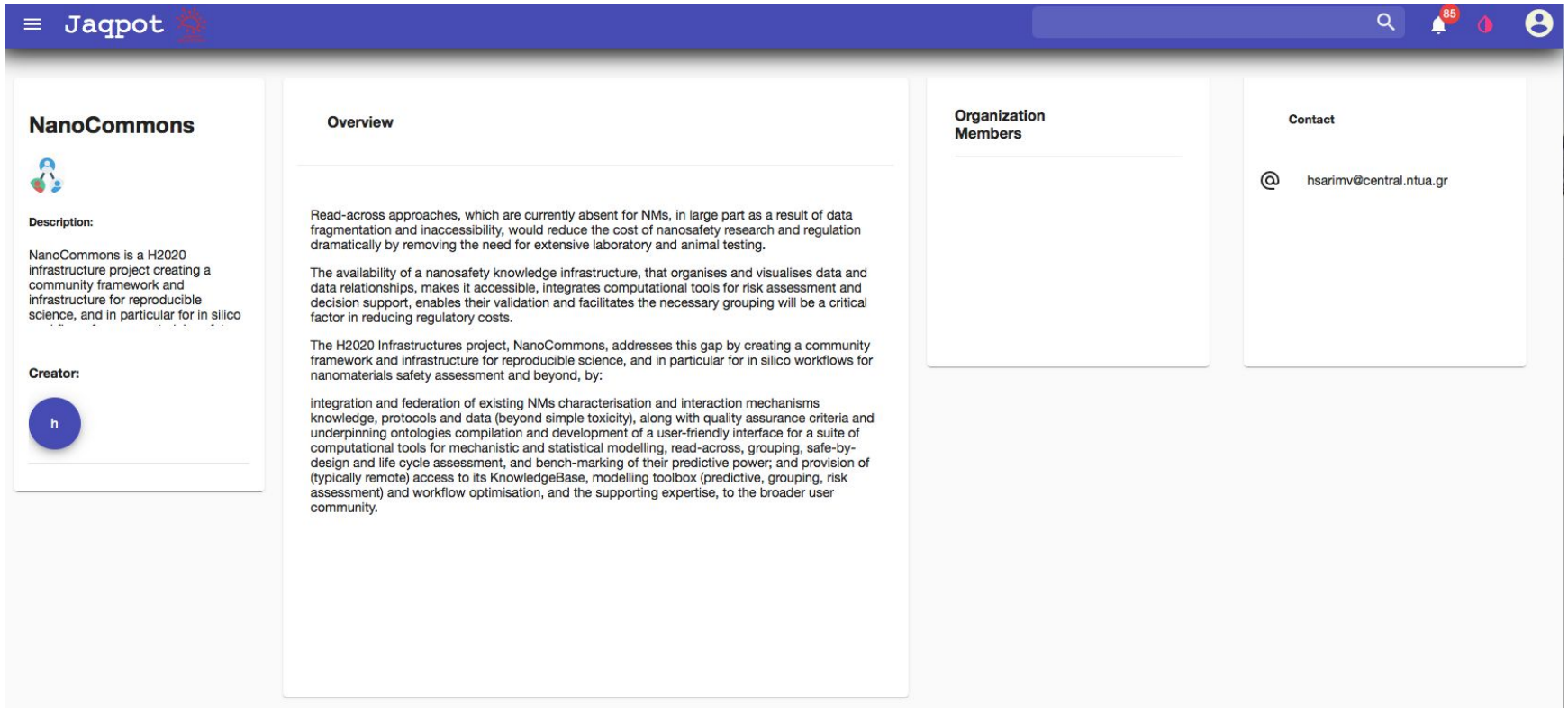


Jaqpot 5 - Organizations



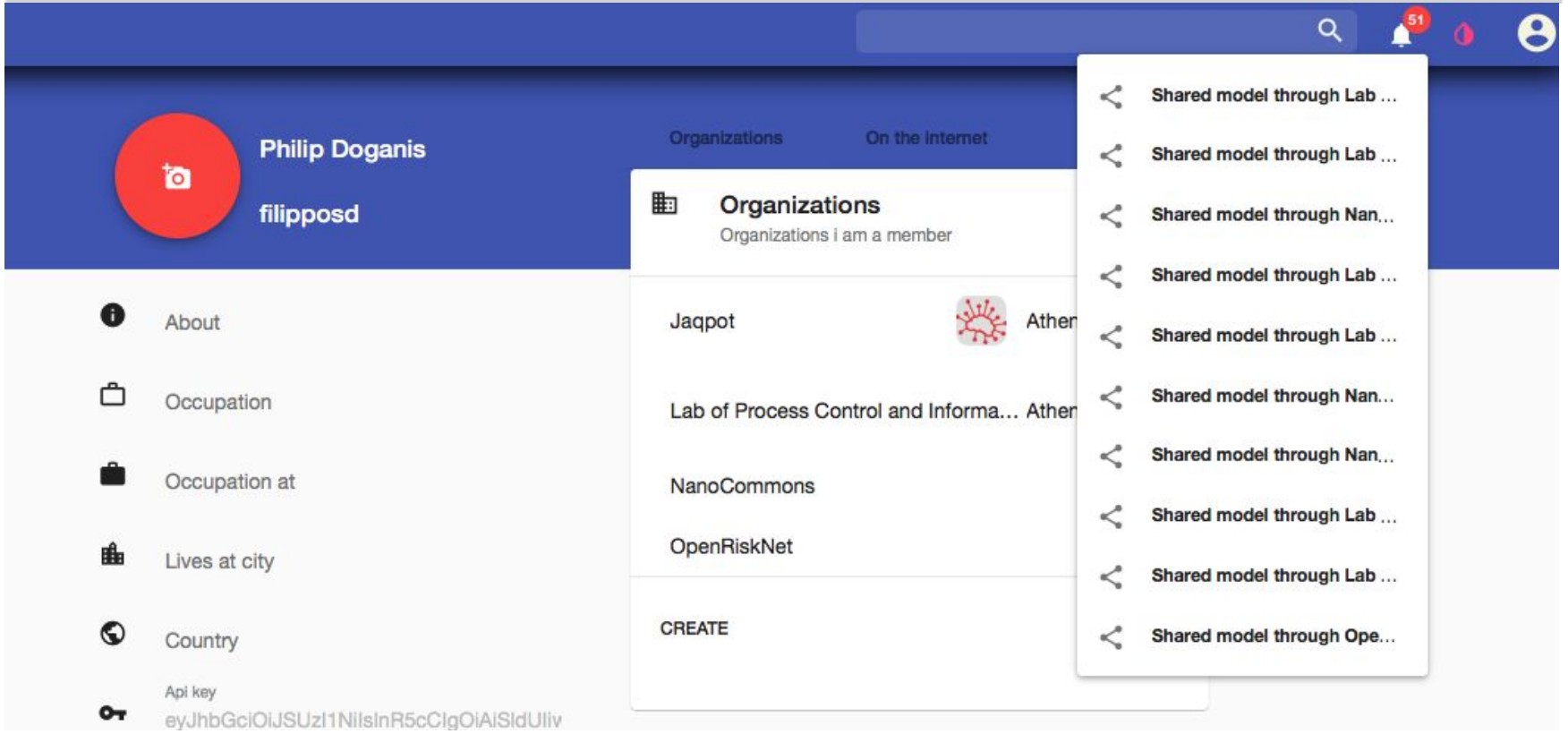
https://app.jaqpot.org/organization/NanoCommons

Jaqpot 5 - NanoCommons Organization page

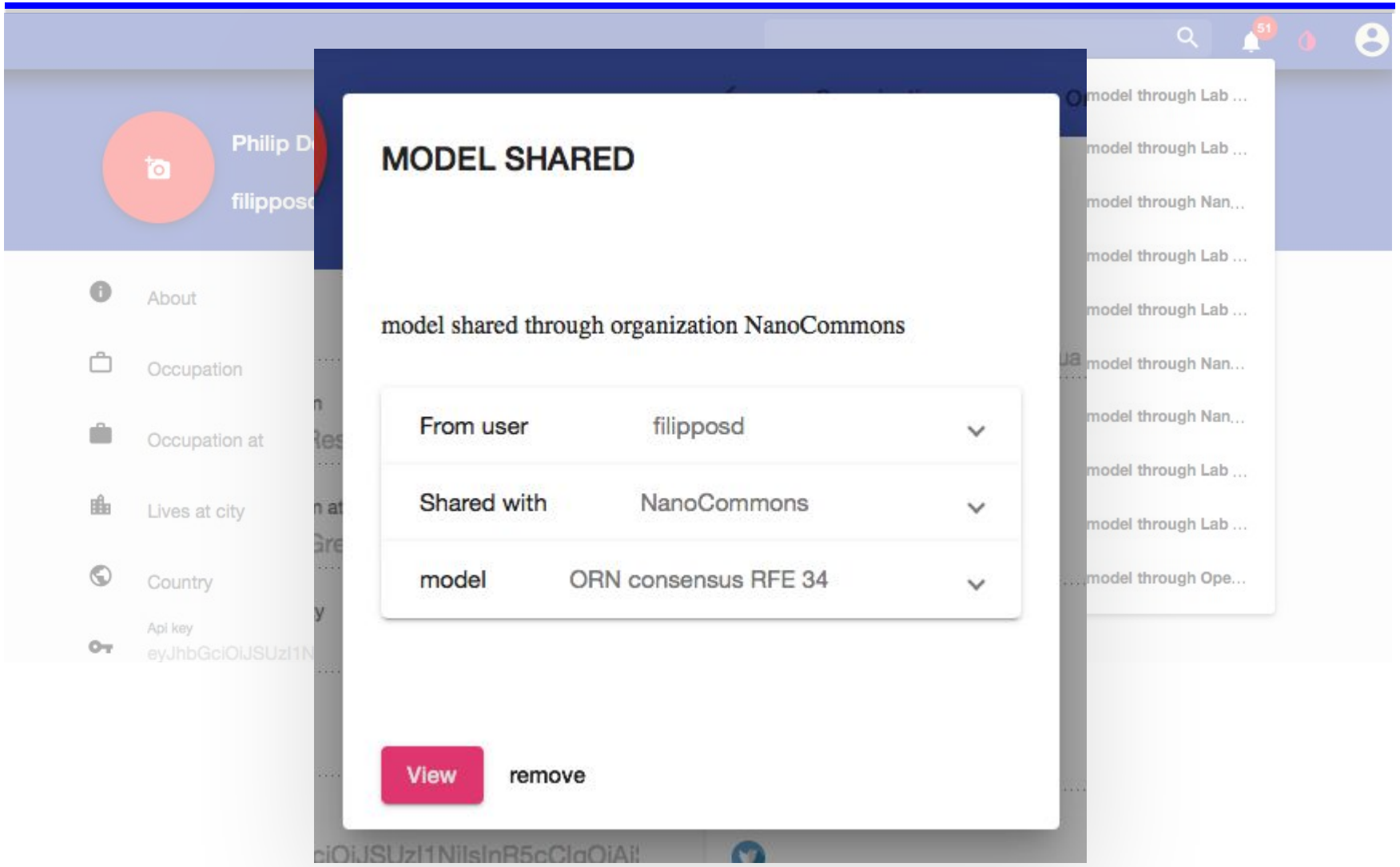


https://app.jaqpot.org/organization/NanoCommons

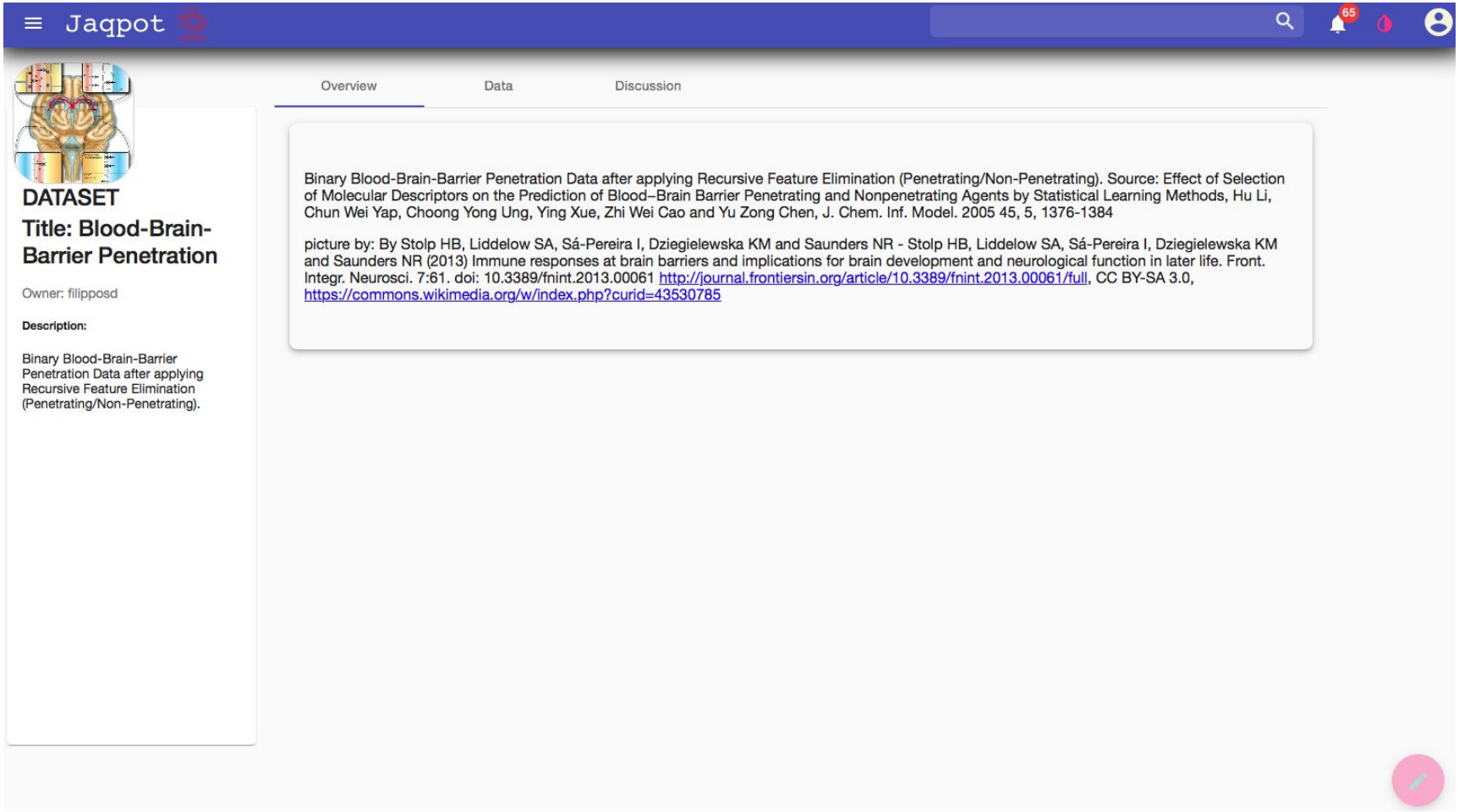
Jaqpot 5 - Sharing of resources



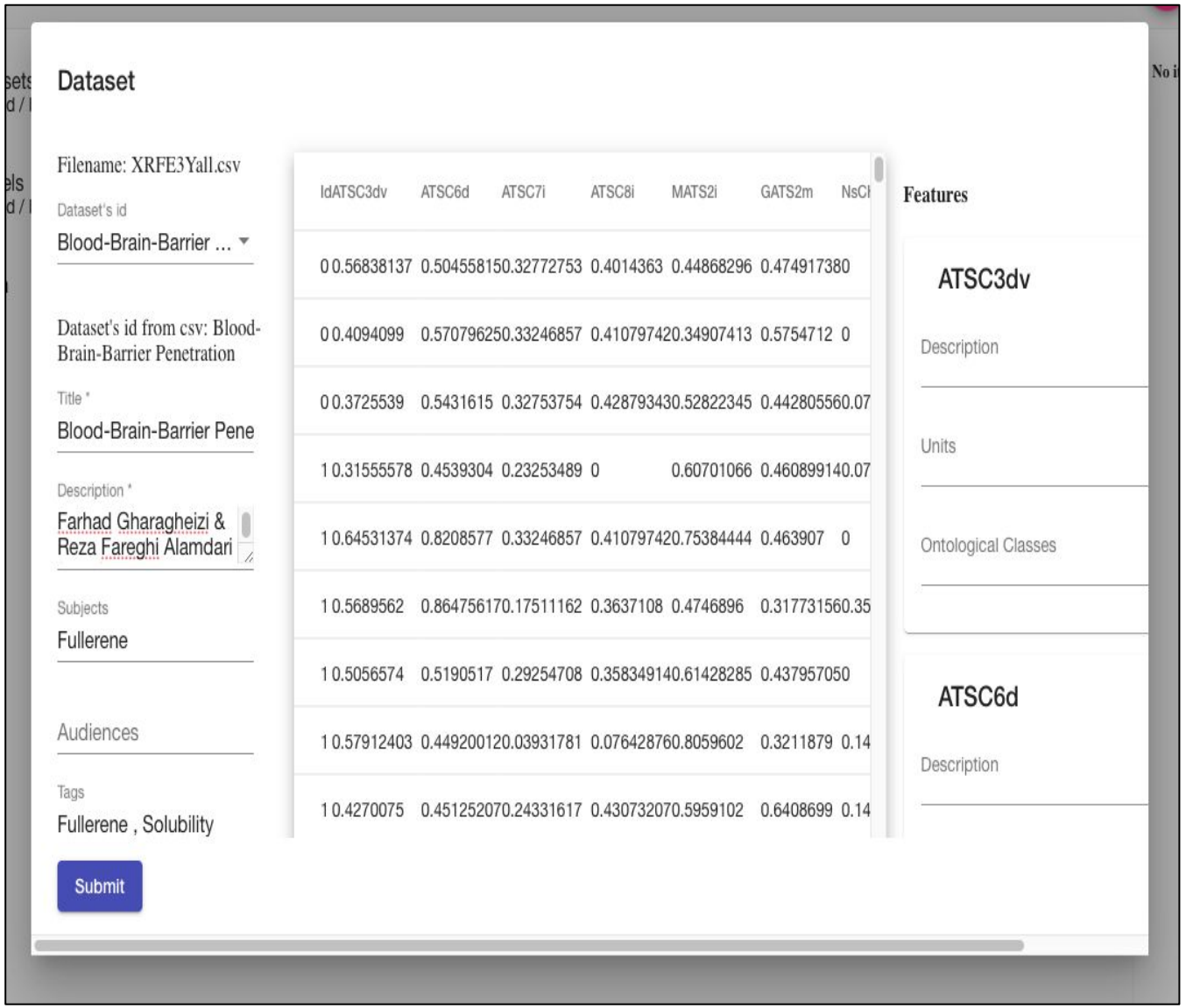
Jaqpot 5 - Sharing notification



Jaqpot 5 - Dataset page



Jaqpot 5 - New dataset -Dataset details



Training a model in Python

```
[ ]  from sklearn.linear_model import LogisticRegressionCV
     clf_RFE2 =LogisticRegressionCV(Cs=100, class_weight=None,
                                    cv=None, dual=False, fit_intercept=True, intercept_scaling=1.0,
                                    max_iter=100,multi_class='ovr', n_jobs=1, penalty='l2',
                                    random_state=None, refit=True, scoring=None,
                                    solver='liblinear', tol=0.0001, verbose=0).fit(XRFE3,Yall1)
     clf_RFE2.predict(XRFE3)
     clf_RFE2.score(XRFE3,Yall1)
```

Now you have a model on your computer.

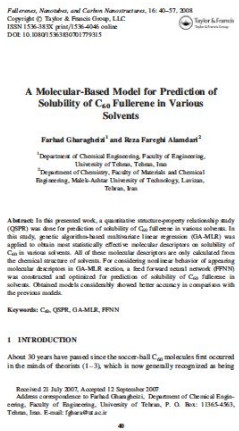
Great.

From model to web service in 1 line

```
url=jaqpot.deploy_pipeline(clf_RFE2,XRFE3,Yall1,"ORN consensus RFE 6","Logistic Regression+RFE","linearmodel")
url
```

jaqpot.deploy_pipeline() parameters are:

- pipeline : sklearn pipeline model is a trained sklearn model
- X : pandas dataframe (X variables).
- y : pandas dataframe (y variables).
- title
- description
- algorithm



Fullerenes, Nanotubes, and Carbon Nanostructures, 16: 40–57, 2008
Copyright © Taylor & Francis Group, LLC
ISSN 1536-383X print/1536-4046 online
DOI: 10.1080/15363830701779315

A Molecular-Based Model for Prediction of Solubility of C_{60} Fullerene in Various Solvents

Farhad Gharagheizi[1] and Reza Fareghi Alamdari[2]

[1]Department of Chemical Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran
[2]Department of Chemistry, Faculty of Materials and Chemical Engineering, Malek-Ashtar University of Technology, Lavizan, Tehran, Iran

Abstract: In this presented work, a quantitative structure-property relationship study (QSPR) was done for prediction of solubility of C_{60} fullerene in various solvents. In this study, genetic algorithm-based multivariate linear regression (GA-MLR) was applied to obtain most statistically effective molecular descriptors on solubility of C_{60} in various solvents. All of these molecular descriptors are only calculated from the chemical structure of solvents. For considering nonlinear behavior of appearing molecular description in GA-MLR section, a feed forward neural network (FFNN) was constructed and optimized for prediction of solubility of C_{60} fullerene in solvents. Obtained models considerably showed better accuracy in comparation with the previous models.

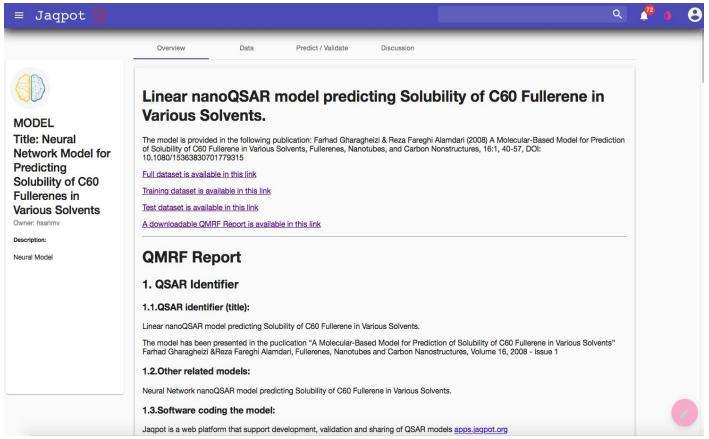
Keywords: C_{60}, QSPR, GA-MLR, FFNN

1 INTRODUCTION

About 30 years have passed since the soccer-ball C_{60} molecules first occurred in the minds of theorists (1–3), which is now generally recognized as being

Received 21 July 2007; Accepted 12 September 2007
Address correspondence to Farhad Gharagheizi, Department of Chemical Engineering, Faculty of Engineering, University of Tehran, P. O. Box: 11365-4563, Tehran, Iran. E-mail: fgharagh@ut.ac.ir

```
url=jaqpot.deploy_pipeline(clf_RFE2,XRFE3,Yall1,"ORN consensus RFE 6","Logistic Regression+RFE","linearmodel")
url
```



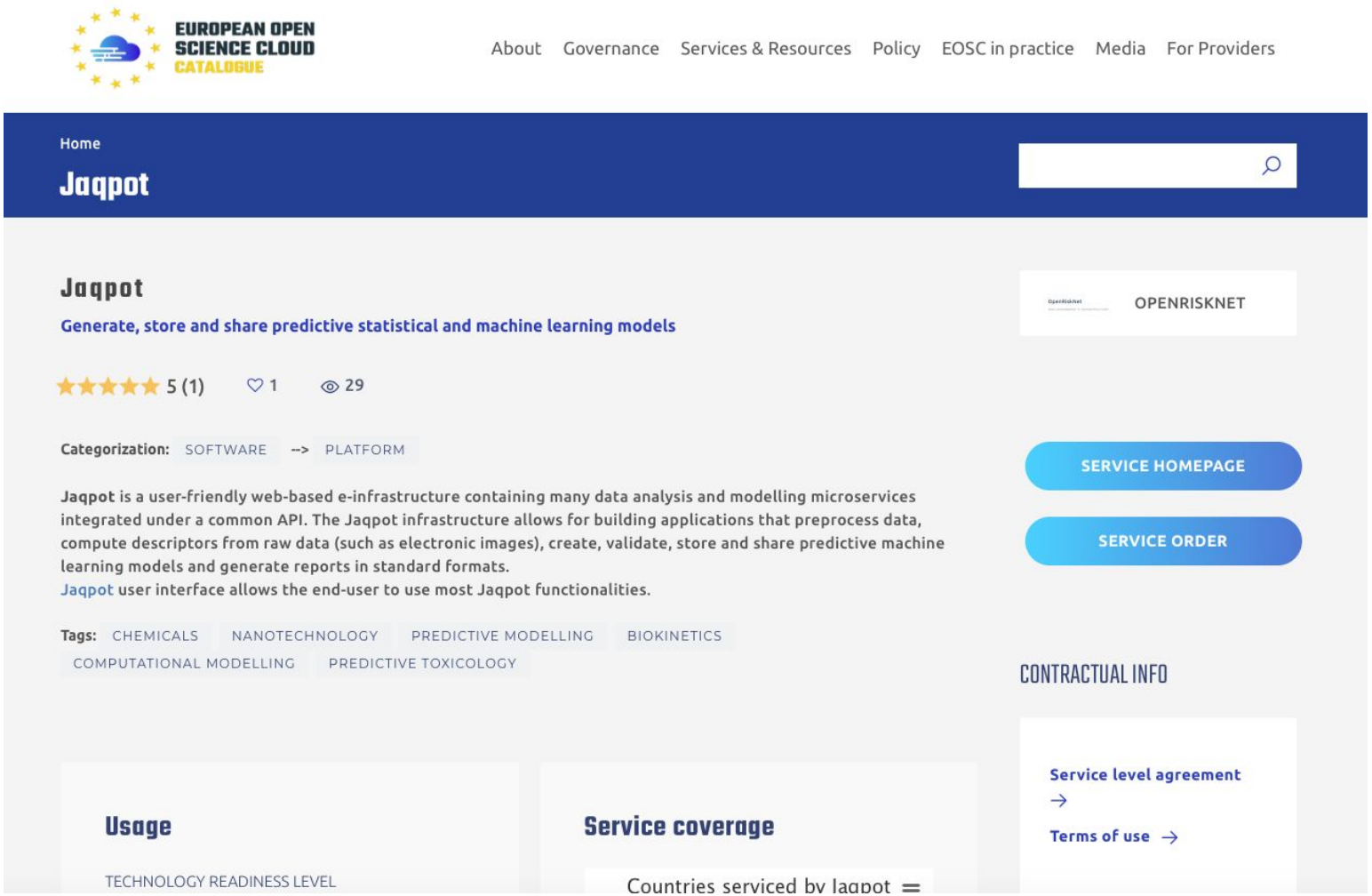
Get predictions from a model

```
[ ]  dfJQ_RFE, predicts_RFE = jaqpot.predict(XRFE3, modelId=url)
```

```
[ ]  dfJQ_RFE
```

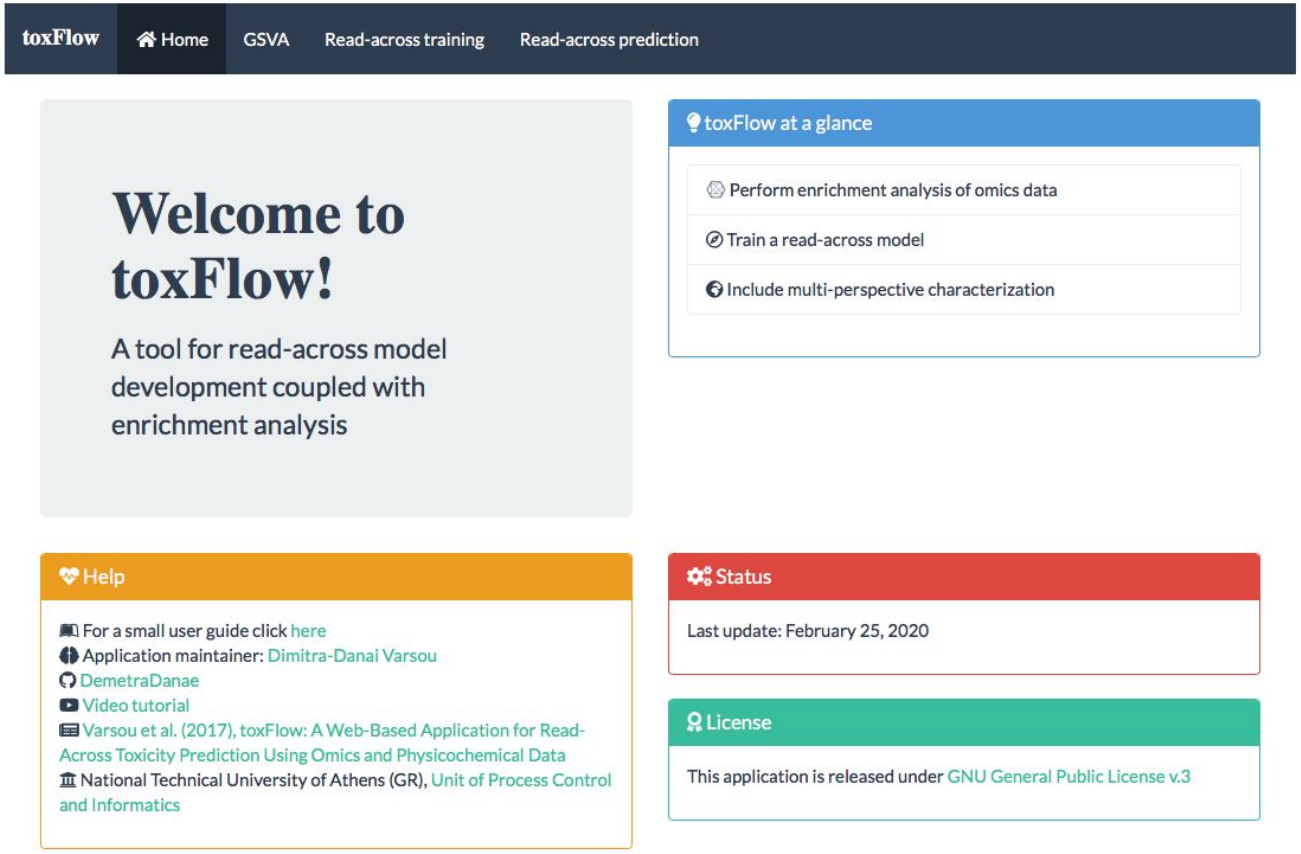
	NsCH3	PEOE_VSA4	ATSC7i	NdssS	ATSC3dv	JGI5	SsCH3	GATS2m	Blood-Brain-Barrier Penetration	ATSC6d
0	0.000000	0.171888	0.327728	0	0.568381	0.207983	0.000000	0.474917	1	0.504558

Jaqpot in EOSC Catalogue (through OpenRiskNet)



https://catalogue.eosc-portal.eu/service/openrisknet.jaqpot

toxFlow



A tool for read-across model development coupled with enrichment analysis

- Perform enrichment analysis of omics data
- Train a read-across model
- Include multi-perspective characterization

toxflow.jaqpot.org

Conclusion

- Jaqpot is an **open source web infrastructure** with flexibility:
  - **generalization across disciplines**
  - **integration to 3rd-party tools** (modelling on the Cloud)
  - **easy integration** to diverse **architectures/ontologies/knowledge domains**.
- Jaqpot offers a seamless way to:
  - **take models out of the desktop or paper** and automatically make them available as web services via their URI
  - functionality is provided both through an **API and GUIs**.

Jaqpot constitutes a universal platform
that exchanges and produces semantically annotated datasets
and employs machine learning techniques to generate predictive models.

A place to

create discuss

share use

your models & datasets

NanoCommons

Nano-Knowledge Community

Thank you

Any Questions?

Jaqpot hands-on session - Meet the team

Philip
Doganis

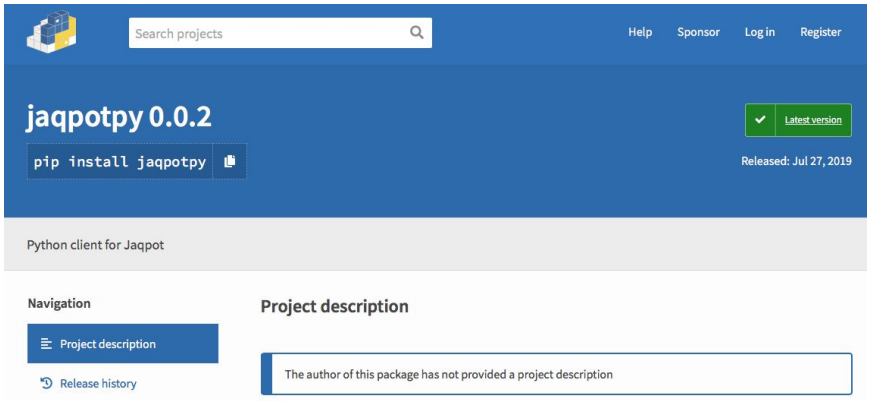
Irini
Liampa

Pantelis
Karatzas

Jupyter: Setting up + brief intro

- Jupyter: Setting up + brief intro to what you need to get started in python
- After anaconda setup
- Environments
- Libraries
- import Jaqpot over pip

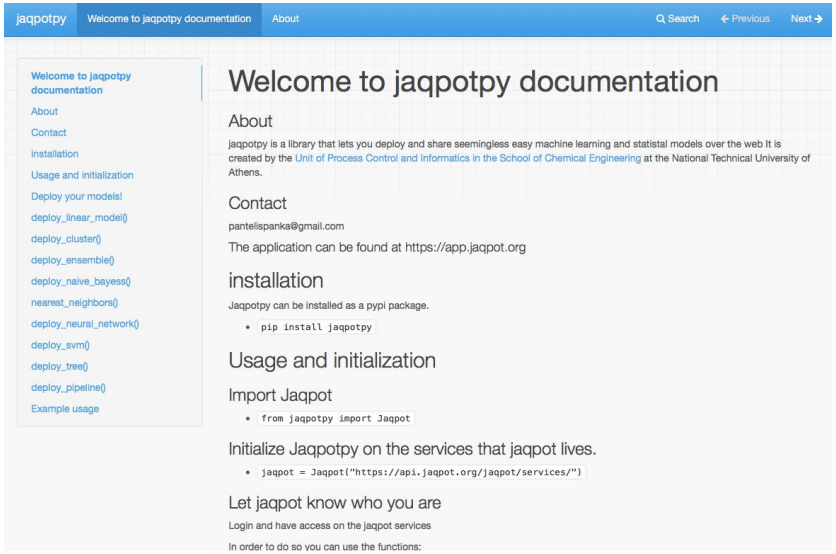
Jupyter: Setting up + brief intro



Jaqpotpy can be installed as a pypi package.

https://pypi.org/project/jaqpotpy/

```
pip install jaqpotpy
```



https://jaqpotpy.readthedocs.io

Importing and exploring the dataset

- Libraries and command for import
- Exploring the features
- Dataframes etc.

Basic preprocessing

- Basic preprocessing
- Checking the dataset
- Performing scaling

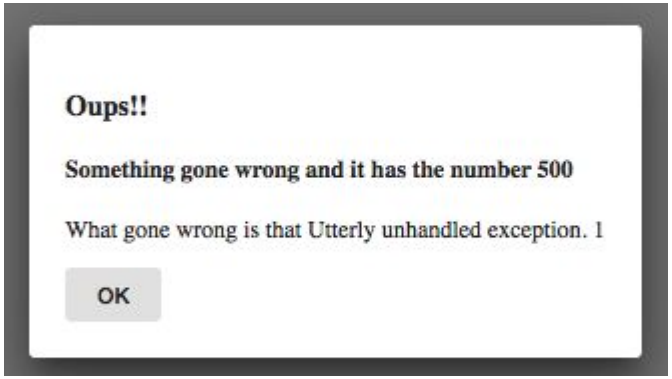
Creating a local model and its evaluation

- Creating a local model and its evaluation
- Create a model
- Evaluate performance
- Graphs and metrics

Uploading your model to Jaqpot

You now have your web service!

- Exploring your model on Jaqpot
  - adding information
  - QMRF report
    - To assist the users in adding editable versions of QMRF reports, we have created a QMRF markdown template, which can be downloaded from the following address: https://github.com/ntua-unit-of-control-and-informatics/QSAR-Models/blob/master/QMRF%20template.md. The user only needs to provide the necessary information under each section and the QMRF report is generated in an easy-to-read format.
- Discussions



Using the model to get predictions on the web and in Python

- Using model on UI
- Using model over API in Python

Explore the platform, create more models and give us your ideas

- Feedback questionnaire
- Jaqpot discussions