

Hands-on Workshop on Computational Nanotoxicology using eNanoMapper web services by NTUA

Workshop:	NMP project eNanoMapper
DATE / PLACE:	March 17, 2016 / National Technical University of Athens, <u>Zografou Campus</u> :
TIME:	13.30

TITLE:	Developing Nano-QSAR predictive toxicology models
---------------	---

SPEAKER:	Philip Doganis and Haralambos Sarimveis
AUTHORS:	Philip Doganis, Haralambos Sarimveis <i>School of Chemical Engineering, National Technical University of Athens</i>

ABSTRACT

The workshop will offer hands-on work on the development of nanoQSAR models based on data available from the data.enanomapper.net server, making use of the eNanoMapper computational infrastructure from NTUA¹ that extends the OpenTox API². The focus is on the use case of predicting cell association of Gold Nanoparticles, based on experimental data, publically available in the publication of Walkey et. al.³.

Participants will go through the workflow of constructing a dataset from a selected one from the eNanoMapper database, performing preparatory operations on it, such as selection of a subset of properties or calculated descriptors and examine using diverse sets of nanoparticle descriptors, including descriptors derived from protein-corona information. Three model building cases will be explored towards the creation of predictive models using statistical and machine learning algorithms, firstly on physicochemical data, secondly on proteomics data and finally a case where the use of PMML (Predictive Model Markup Language) in XML-based files for definitions of descriptor transformations and model output will be presented based on physicochemical data.

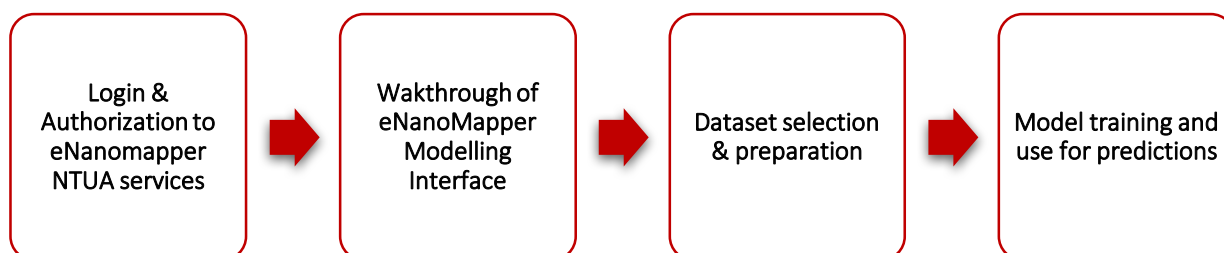


Figure 1 Workshop Outline

1. Login & Authorization to the Jaqpot and eNanoMapper services

Users access the Jaqpot Quattro homepage by NTUA <http://app.jaqpot.org:8000/> (Figure 2), currently as a test instance as is under development. To gain access to the services, users should click the **Sign in** button, to be transferred to the screen in Figure 3.

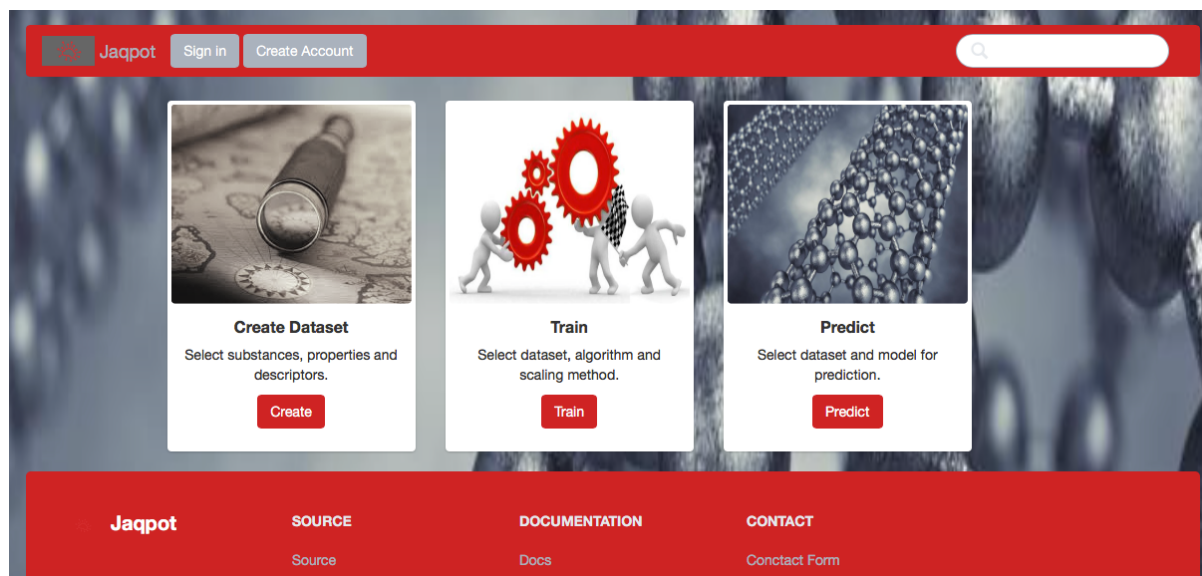


Figure 2 Homepage of eNanoMapper NTUA Modelling interface

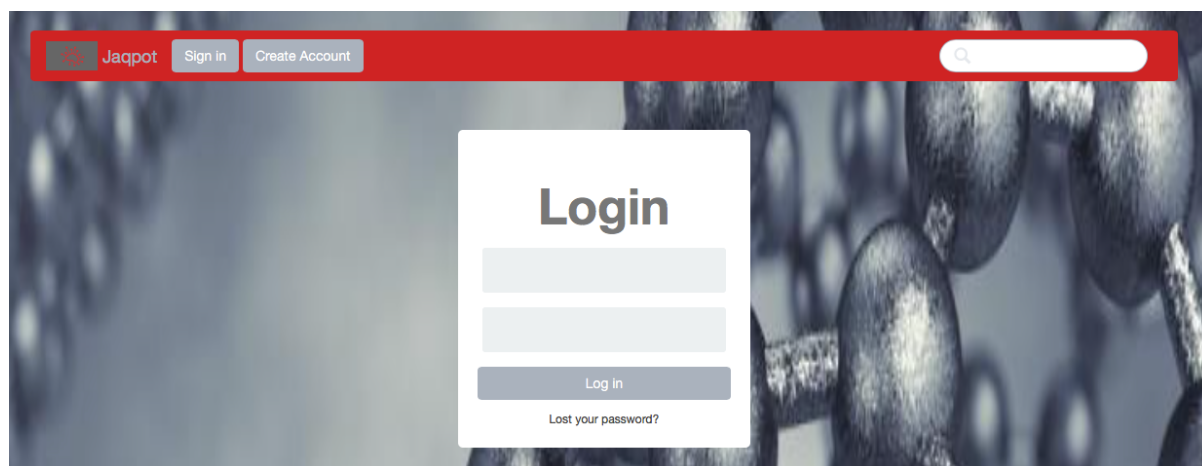


Figure 3 Login page

Log in to Jaqpot Quattro using OpenTox credentials. You can get an OpenTox user account by clicking on “Create Account” on the start page (Figure 2) at: http://opentox.org/join_form (Figure 4), login with your own credentials or just type “guest” in both fields. Successful entry of credentials leads to the homepage for registered users (Figure 5).

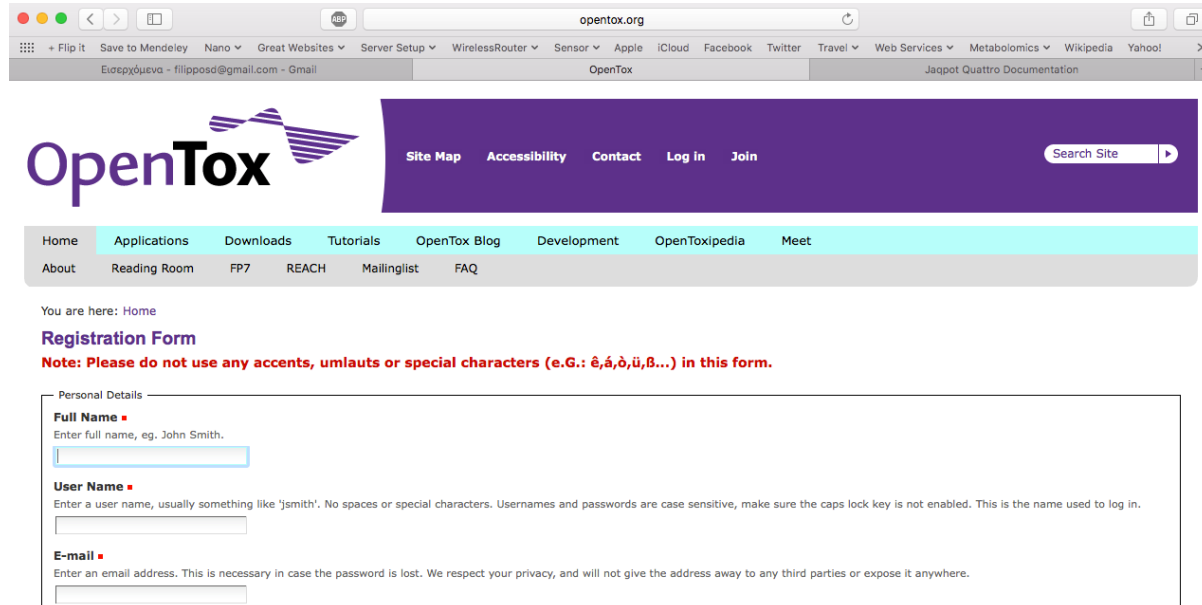



Figure 4 OpenTox Registration page (http://www.opentox.org/join_form)

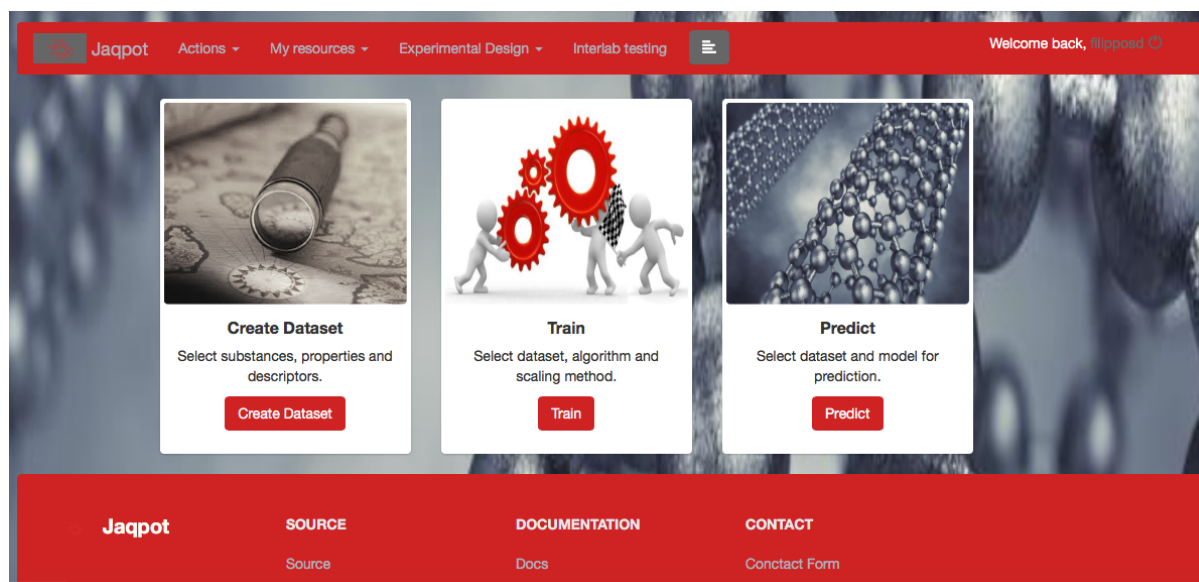


Figure 5 Starting page for registered users

2. eNanoMapper Jaqpot modelling services – A walkthrough

In this section we will have a brief walkthrough of the eNanoMapper Jaqpot modelling interface. Please note that this is a test User interface for the Modelling API (Application programming interface), which can be examined by more proficient users who would like to use the Web Services exposed through it in order to achieve more automated modelling workflow. Users can visit <http://app.jaqpot.org:8080/jaqpot/swagger/> to view the documentation of the API and experiment with its possibilities using the Swagger interface hosted there (Figure 6). As the eNanoMapper Jaqpot

modelling interface leaves the testing phase and reaches finalization, a manual for its functionality will be released.

Jaqpote Quattro

<http://app.jaqpote.org:8080/jaqpote/services/api-docs>

AQIC5wM2LY4SfczCPozaU8C

Explore

report : Report API	Show/Hide	List Operations	Expand Operations	Raw
dataset : Dataset API	Show/Hide	List Operations	Expand Operations	Raw
interlab : Interlab Testing API	Show/Hide	List Operations	Expand Operations	Raw
pmml : PMML API	Show/Hide	List Operations	Expand Operations	Raw
bibtex : BibTeX API	Show/Hide	List Operations	Expand Operations	Raw
validation : Validation API	Show/Hide	List Operations	Expand Operations	Raw
enm : eNM API	Show/Hide	List Operations	Expand Operations	Raw
model : Models API	Show/Hide	List Operations	Expand Operations	Raw
task : Tasks API	Show/Hide	List Operations	Expand Operations	Raw
algorithm : Algorithms API	Show/Hide	List Operations	Expand Operations	Raw
aa : AA API	Show/Hide	List Operations	Expand Operations	Raw
feature : Feature API	Show/Hide	List Operations	Expand Operations	Raw
user : Users API	Show/Hide	List Operations	Expand Operations	Raw

[BASE URL: <http://app.jaqpote.org:8080/jaqpote/services/api-docs>]

Figure 6 eNanoMapper Modelling API: Screenshot of Swagger interface for API documentation

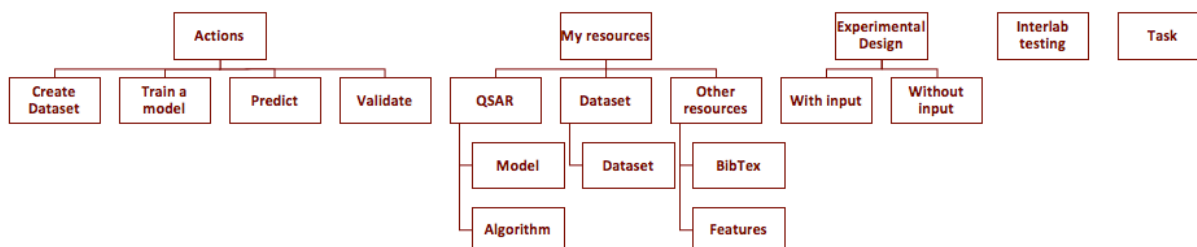


Figure 8 Jaqpote menu tree

The menu tree of the Jaqpote modelling interface is shown in Figure 87. Only the items of interest to this workshop will be described; please also note that this is test release and tools are under development.

The modelling Actions are (see Figure 78):

- **Dataset Creation** from an existing dataset in the data.enanomapper.net database, allowing selection a subset of substances or properties and execution of descriptor calculations when image files or

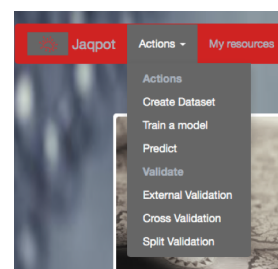


Figure 7 Available modelling actions

Crystallographic data are included in the dataset, resulting to the creation of a local dataset.

- **Model training** using a dataset in the data.enanomapper.net database or the local dataset and applying one of the available algorithms, currently:
 - Regression
 - *ocpu-lm*: Linear Regression by R offered through OpenCPU
 - *weka-mlr*: Linear Regression by Weka
 - *weka-pls*: Partial Least Squares by Weka
 - *weka-svm*: Support Vector Machines by Weka
 - *python-pls-vip*: Partial Least Squares with “Variable Importance in Projection” scores
 - Classification
 - *weka-pls*: Partial Least Squares Classification by Weka
 - *python-id3-mci*: ID3 classification algorithm

Users can export their model in PMML format (explained in following chapter) and distribute it online using the unique URI assigned to it for easy deployment to other systems/web services. More detailed descriptions on the algorithms, their implementation and their parameters can be found in the openly available Deliverable 4.3 of eNanoMapper.

- **Predict** by applying a model to a dataset with matching properties and receive a dataset with predictions and Domain of Applicability calculations, if that choice has been made.

The resources that are available to the user are:

- **QSAR**
 - **Model**: a repository of models created by the user and public models
 - **Algorithm**: a list of the available algorithms offered by Jaqpot
- **Dataset**
 - A repository of datasets created by the user and public models

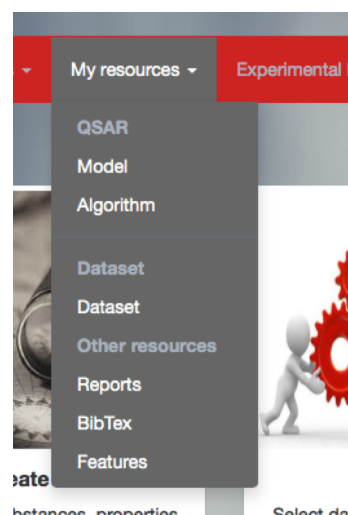


Figure 9 “My resources” menu

3. Dataset selection & preparation

Three cases will be examined, all involving the data from Walkey et.al.³. The **first** case involves a dataset with only Physicochemical data and two toxicity related endpoints, Net Cell Association and its log2 value (only one will be selected for use in the next). The **second** case involves a dataset with only Proteomics data and a toxicity related endpoint. The **third** case involves using PMML input files for transformations and determination of input variables.

Case 1

In the first case, a dataset with 84 Gold nanoparticles will be used after having been split into two sets, a Training and a Prediction set with 56 and 28 nanoparticles respectively. Among the available properties, only the Physicochemical data will be selected, thus leaving out the Proteomics experimental data, as shown in Figure 10.

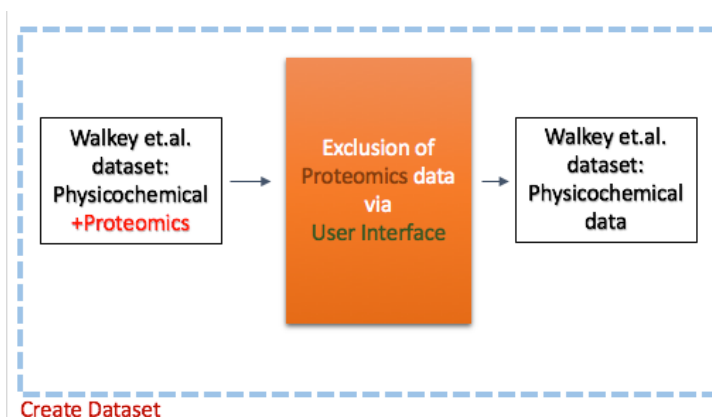


Figure 10 Create Dataset actions for Case 1

After clicking the “Create Dataset” option in the start screen (Figure 5) a view of datasets available on the <https://apps.ideaconsult.net/enmtest/ui> database is shown. Please select the **WalkeyGoldTrain.csv** file and hit Next, as in Figure 11.

Create Dataset

Choose method:

☒ Select Substance owner.
Complete substance owner id.

Select substance owner:

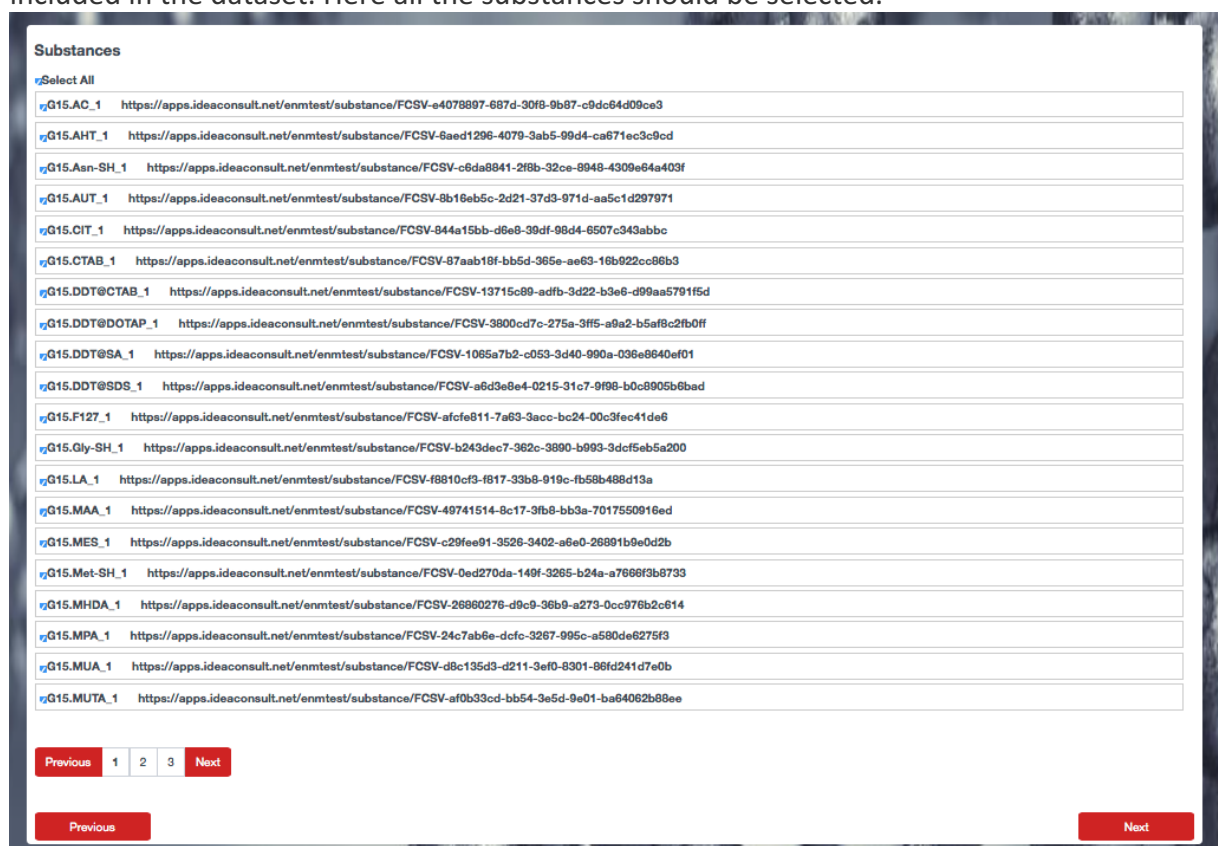
Name	Subcategory	Count
FCSV-A9367169-11DF-3AA8-9A7A-DCDA6893F2A4	publishProperty.csv	1
FCSV-ACC0E543-2CE0-3B78-B412-8E03FCD1F5BC	ProteinCorona11.csv	1
FCSV-ACCFB01A-0D34-38D1-BEA3-D1AEB5AC21A7	PathakotiCSVdos1.csv	1
FCSV-B0E5A5AC-FD96-3B36-81F8-242CA92643DF	WalkeyGoldTrain.csv	56
FCSV-B67B0E30-133E-3182-BB11-5C0C4C4C3C8B	ProteinCorona_4_prediction_1398219745.csv	84
FCSV-B8A9C515-7A79-32A9-83D2-8FF2FEC8ADCB	ProteinCorona_5.csv	84
FCSV-C8C319A7-6114-3124-AD4D-75AAD4FE917A	PathakotiCSV.csv	1
FCSV-CA78016F-6091-39C7-B0DF-E52037736624	ProteinCorona_4.csv	84
FCSV-D252A291-36C9-307E-86DC-F59B398A50F2	PathakotiCSVdos.csv	1
FCSV-E22C3085-2A7F-3E22-88AF-8DC273722708	GOdescr_train_onlyImpGO_24groups1.csv	63
FCSV-E64468E1-4A7C-3570-B73F-D2D42CB7BDDF	WalkeyGoldPredict.csv	28
FCSV-EE25FA92-B366-39CA-8767-C25F3C0DD88C	MergedSheetsGoldNoNeutralsClean.csv	84
FCSV-F1012C60-26DE-356A-BD1E-28E3DEF04B28	import3mtox.csv	1
FCSV-F56AD293-2033-334F-A323-CE507C023702	PuzynDataCSV.csv	3
FCSV-F9B1B49D-73F4-300A-BF6D-9B1ED7E9B148	ProteinCorona_8V2_prediction_1674942482.csv	84
FCSV-F9F67F08-D4A9-3B19-9419-6E2F556DE00D	PuzynDatasetWorkingTrainAsInPaperCorrect.csv	10
FCSV-FE19E6AF-4A4C-3CDF-8C3E-08BE07237ABA	ProteinCorona_8V2.csv	84
IUC4-44BF02D8-47C5-385D-B203-9A8F315911CB	OECD / Paris / France	3
IUC5-354430C7-746E-450E-9722-09788EB90A29	Ideaconsult Ltd. / Sofia / Bulgaria	1
NWKI-9F4E86D0-C85D-3E83-8249-A858659087DA	NanoWiki	337

Previous ... Next

Next

Figure 11 Create dataset: Selection of substance owner

In the **Substances** screen (Figure 12) users can select a subset of the available substances to be included in the dataset. Here all the substances should be selected.



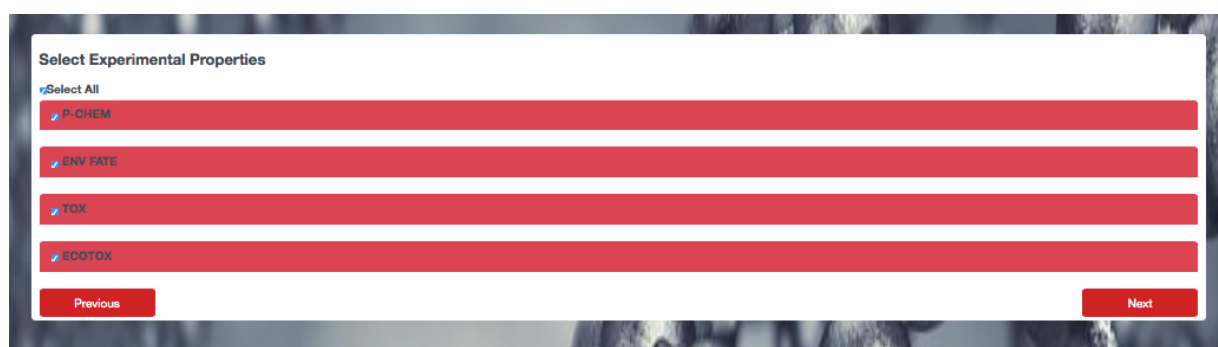
Substances	
☒ Select All	
☒ G15.AC_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-e4078897-687d-30f8-9b87-c9dc84d00ce3
☒ G15.AHT_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-6aed1296-4079-3ab5-99d4-ca671ec3c9cd
☒ G15.Asn-SH_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-c8da8841-2f8b-32ce-8948-4309e84e403f
☒ G15.AUT_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-8b16eb5c-2d21-37d3-971d-aa5c1d297971
☒ G15.CIT_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-944a15bb-d6e8-30df-98d4-6507c343abb0
☒ G15.CTAB_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-87aab18f-bb5d-365e-ae63-16b922cc86b3
☒ G15.DDT@CTAB_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-13715c89-adfb-3d22-b3e6-d99aa5791f5d
☒ G15.DDT@DOTAP_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-3800cd7c-275a-3ff5-a0a2-b5af8c2fb0ff
☒ G15.DDT@SA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-1065a7b2-c053-3d40-990a-036e8640ef01
☒ G15.DDT@SDS_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-a6d3e8e4-0215-31c7-9f98-b0c8905b6bad
☒ G15.F127_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-afcf811-7a63-3acc-bc24-00c3fec41de6
☒ G15.Gly-SH_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-b243dec7-362c-3890-b993-3dcf5eb5a200
☒ G15.LA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-f8810cf3-f817-33b8-919c-fb58b488d13a
☒ G15.MAA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-49741514-8c17-3fb8-bb3a-7017550916ed
☒ G15.MES_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-c29fee91-3528-3402-a6e0-26891b9e0d2b
☒ G15.Met-SH_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-0ed270da-149f-3265-b24a-a7686f3b8733
☒ G15.MHDA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-26860276-d9c9-36b9-a273-0cc976b2c614
☒ G15.MPA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-24c7ab8e-dcfc-3267-995c-a580de6275f3
☒ G15.MUA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-d8c135d3-d211-3ef0-8301-86fd241d7e0b
☒ G15.MUTA_1	https://apps.ideaconsult.net/enmtest/substance/FCSV-af0b33cd-bb54-3e5d-9e01-ba64062b88ee

Previous 1 2 3 Next

Previous Next

Figure 12 Create dataset: Substance select page

In the next screen (Figure 13), users may select the experimental properties to be included in the dataset, which have been grouped into 4 categories, Physicochemical, Environmental Fate, Toxicity and Ecotoxicity.



Select Experimental Properties

☒ Select All

☒ P-CHEM

☒ ENV FATE

☒ TOX

☒ ECOTOX

Previous Next

Figure 13 Create dataset: Selection of experimental properties to be included in the dataset

In the case studied, the Proteomics section in the Toxicity Category should be unchecked as in Figure 14, as the scope of the case is to study the Physicochemical section.

☒ TOX

☒ TO_ACUTE_ORAL_SECTION
☒ TO_ACUTE_INHAL_SECTION
☒ TO_ACUTE_DERMAL_SECTION
☒ TO_SKIN_IRRITATION_SECTION
☒ TO_EYE_IRRITATION_SECTION
☒ TO_SENSITIZATION_SECTION
☒ TO_SENSITIZATION_HUMAN_SECTION
☒ TO_SENSITIZATION_INVITRO_SECTION
☒ TO_SENSITIZATION_INCHEMICO_SECTION
☒ TO_REPEATED_ORAL_SECTION
☒ TO_REPEATED_INHAL_SECTION
☒ TO_REPEATED_DERMAL_SECTION
☒ TO_GENETIC_IN_VITRO_SECTION
☒ TO_GENETIC_IN_VIVO_SECTION
☒ TO_CARCIINOGENICITY_SECTION
☒ TO_REPRODUCTION_SECTION
☒ TO_DEVELOPMENTAL_SECTION
☒ UNKNOWN_TOXICITY_SECTION
☒ PUBCHEM_CONFIRMATORY_SECTION
☒ PUBCHEM_SUMMARY_SECTION
☒ PUBCHEM_SCREENING_SECTION
☒ PUBCHEM_DOSERESPONSE_SECTION
☒ PUBCHEM_PANEL_SECTION
☐ PROTEOMICS_SECTION
☒ BAO_0003009_SECTION
☒ BAO_0002993_SECTION
☒ BAO_0002100_SECTION
☒ BAO_0002167_SECTION
☒ BAO_0002167_SECTION

Figure 14 Create dataset: Some of the Toxicity properties available for selection. Please notice that the Proteomics Section is unchecked.

In the next stage (Figure 15), users can select the descriptors to be transferred, in the case of (pre-existing) Experimental values or to be calculated, in case the dataset contains nanoparticle images, Crystallographic data, Proteomics data and descriptors based on CDK or algorithms can be calculated. Additionally, the user must select a representative title and a description for the dataset.

Clicking the Next button at this stage leads to a notification page that the task has been initiated and an intermediate page showing the status of the task until completion (both pages shown in Figure 16).

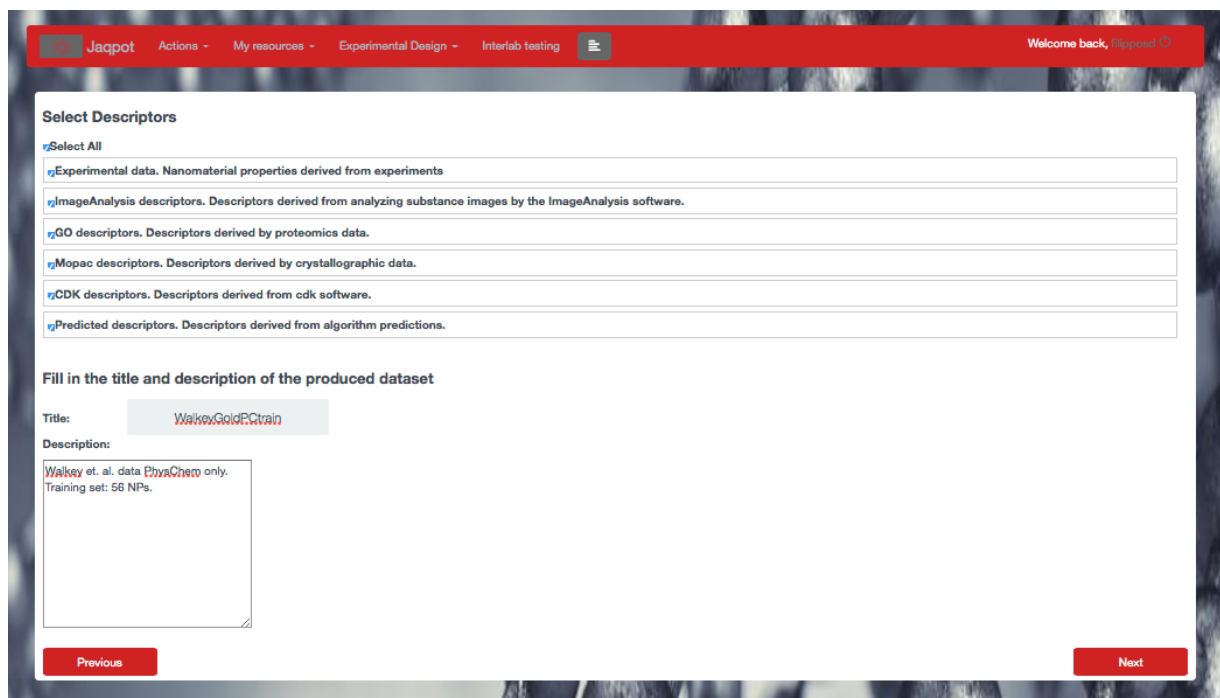


Figure 15 Create dataset: Selection of descriptors to be transferred or calculated, assignment of title and description

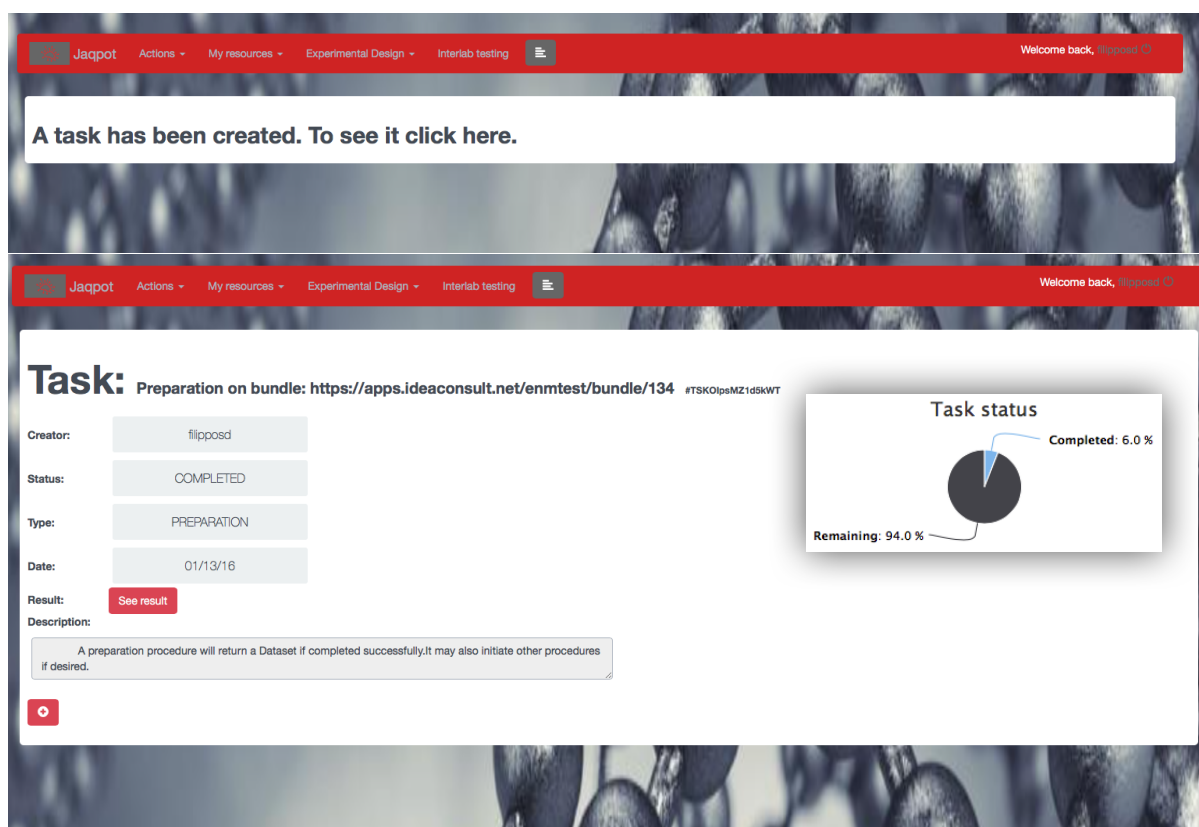


Figure 16 Intermediate task screens: Above: notification of task initiation. Below: Task progress page. While task is running, the status pie chart updates without refresh.

The next page shows the view of the dataset in the Jaqpot modelling interface (Figure 17).

Dataset: WalkeyGoldPCtrain HKL426C8J5mp

Search:

Compounds	Polyspecificity Index	Density	Intensity Mean Nanoparticle Diameter	Absorbance	Z-Average Nanoparticle Diameter	Number Mean Nanoparticle Diameter	Net cell association	LDH5000	Polyspecificity Index	Protein density	ZETA POTENTIAL	Localized Surface Plasmon Resonance LSPR index	Log2 transformed	Local Surface Plasmon LSPR
G15.AC_1	0.084	19.1	70.97	255.443	57.53	23.49	0.028	518.77	0.27	0.0	-21.78	0.454	-5.184	0.183
G15.AHT_1	0.399	19.1	106.7	240.729	90.06	47.5	0.497	526.28	0.215	0.0	15.22	0.526	-1.009	0.458
G15.Aan-SH_1	0.15	19.1	68.92	240.177	37.75	25.04	0.02	518.57	0.207	0.0	-20.27	0.327	-5.676	0.274
G15.AUT_1	0.326	19.1	83.34	244.228	55.98	29.49	0.402	523.23	0.273	1.0	16.35	0.39	-1.316	0.365
G15.CIT_1	0.138	19.1	64.47	272.789	54.03	28.3	0.023	517.88	0.217	0.0	-25.26	0.293	-5.42	0.21
G15.CTAB_1	0.485	19.1	105.1	213.324	59.7	35.99	0.017	518.9	0.371	0.0	-5.01	0.368	-5.862	0.326
G15.DOT@CTAB_1	0.191	19.1	99.43	247.743	48.4	23.23	0.005	519.37	0.308	0.0	19.6	0.325	-7.59	0.275
G15.DOT@DOTAP_1	0.193	19.1	83.76	235.484	47.34	31.58	0.458	522.1	0.237	0.0	13.92	0.297	-1.128	0.276
G15.DOT@SA_1	0.249	19.1	88.81	249.152	59.93	55.31	0.009	534.13	0.367	0.0	-29.37	0.321	-6.804	0.396
G15.DOT@SDS_1	0.302	19.1	142.83	208.318	100.13	35.19	0.005	550.6	0.37	0.0	-21.88	0.36	-7.675	0.465
G15.F127_1	0.216	19.1	72.78	238.252	46.83	36.47	0.024	519.37	0.197	0.0	-8.96	0.245	-5.381	0.176
G15.Gly-SH_1	0.211	19.1	74.16	243.954	55.39	38.3	0.032	518.6	0.236	0.0	-17.77	0.403	-4.975	0.282
G15.LA_1	0.23	19.1	84.19	268.117	48.09	39.75	0.016	519.13	0.201	1.0	-26.48	0.301	-5.964	0.237
G15.MAA_1	0.273	19.1	64.85	261.99	60.99	47.92	0.014	518.4	0.238	1.0	-24.17	0.449	-6.142	0.281
G15.MES_1	0.297	19.1	83.75	258.889	61.88	72.5	0.109	518.3	0.286	0.0	-25.55	0.478	-3.199	0.226
G15.Mel-SH_1	0.167	19.1	72.7	260.336	52.42	28.2	0.016	518.0	0.215	1.0	-22.2	0.282	-5.928	0.216
G15.MHDA_1	0.076	19.1	47.82	266.518	43.62	23.95	0.019	522.77	0.155	0.0	-25.65	0.294	-5.778	0.242
G15.MPA_1	0.284	19.1	70.02	261.392	55.65	39.49	0.024	518.67	0.289	0.0	-27.52	0.451	-5.396	0.304
G15.MUA_1	0.092	19.1	51.1	266.113	45.52	22.4	0.035	521.71	0.166	0.0	-21.13	0.302	-4.847	0.236
G15.MUTA_1	0.156	19.1	139.75	263.69	127.34	30.92	1.081	521.84	0.173	0.0	18.17	0.419	0.113	0.25

Previous 1 2 3 Next

Figure 17 View of the dataset produced

Both the Train and the Predict Dataset have been uploaded already and are available in the Jaqpot modelling interface as Example datasets under the names:

Walkey et al - 56 Gold NPs with description: "56 Gold NPs with 25 PhysChem descriptors"

and

Walkey et al - 28 Gold NPs with description: "28 Gold NPs with 25 PhysChem descriptors".

They can be accessed at <http://app.jaqpot.org:8000/data>.

Case 2

In the second case, a dataset originating from the same publication, having the same 84 Gold nanoparticles as in Case 1, but this time with relative abundance data for 76 proteins will be used, after having been split into two sets, a Training and a Prediction set with 56 and 28 nanoparticles respectively.

Similarly to the previous case, users can follow the Create Dataset route to input the data into the Jaqpot Quattro Modelling Interface. For the sake of brevity, both the Train and the Predict Dataset have been uploaded already and are available in the Jaqpot modelling interface as Example datasets under the names:

Walkey et al - 56 Gold NPs with description: "56 Gold NPs with 76 protein corona descriptors"

and

Walkey et al - 28 Gold NPs with description: "28 Gold NPs with 76 protein corona descriptors".

They can be accessed at <http://app.jaqpot.org:8000/data>.

4. Model training and use for predictions

The workflow that will be followed in the three modelling cases is shown in Figure 18.

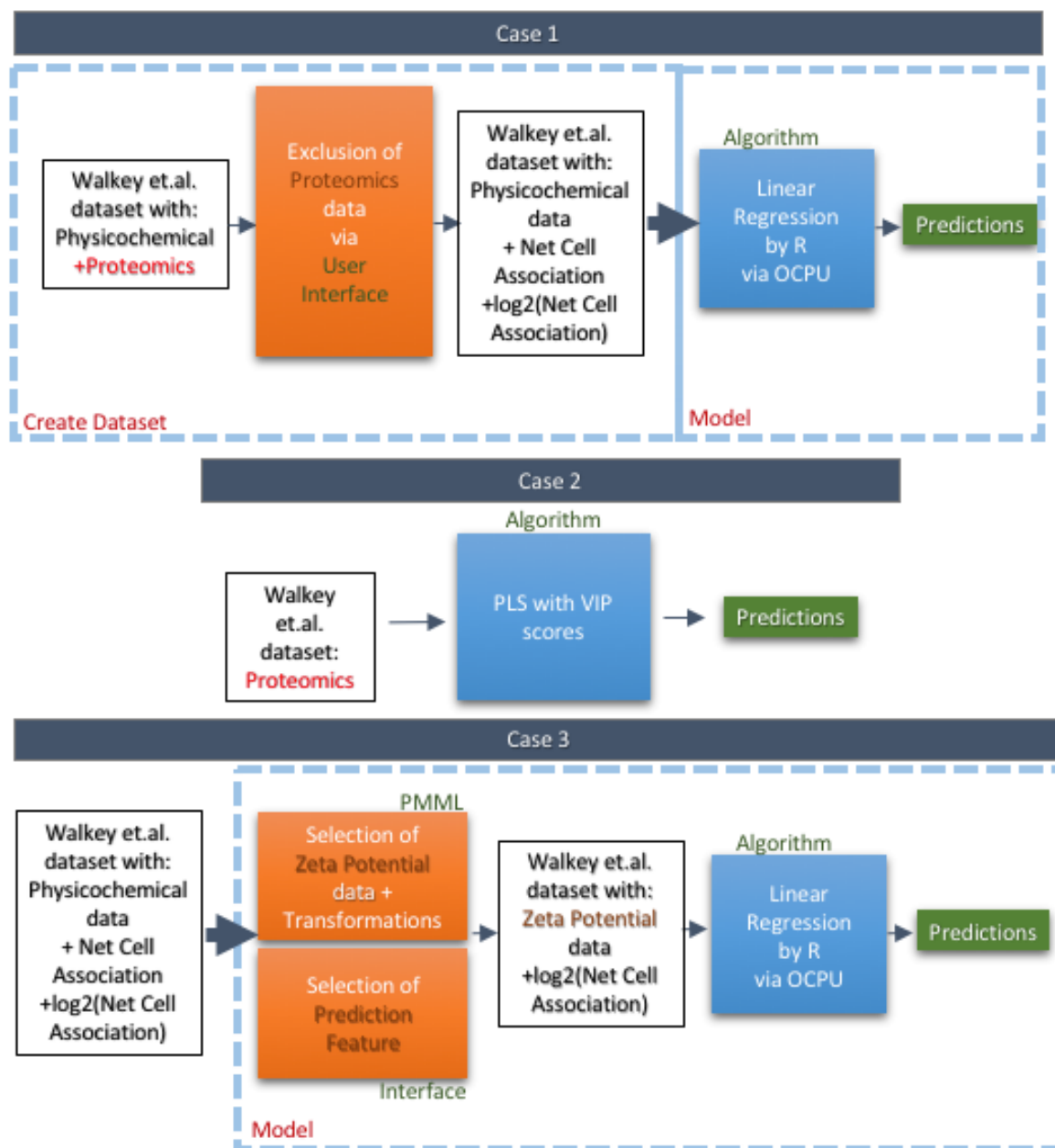


Figure 18 Workflow of the modelling cases

Case 1

Please note that the dataset of Case 1 has undergone processing prior to modelling (dotted “Create dataset” area in figure). The following steps guide the user to performing the modelling operations outlined in Figure 18.

Select dataset:

Example datasets:

Name	Type	Description
PWY25q3v50	Walkey et al - 56 Gold NPs	56 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
KIL42cC8JSmP	Walkey et al - 56 Gold NPs	56 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.
class-dummy	Dummy Classification Dataset	This dataset contains classification data
KEORawkaOng	Walkey et al - 28 Gold NPs	28 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
yzsAXESrLPzz	Walkey et al - 28 Gold NPs	28 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.
UhysSLaF34Spk0	Walkey et al - 84 Gold NPs	84 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
xbR5AMG1r0Bc	Gajewicz et al - 8 Metal Oxide NPs	8 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
E20vuMNNTHap	Gajewicz et al - 10 Metal Oxide NPs	10 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
vNxLUMy33TfM	Gajewicz et al - 18 Metal Oxide NPs	18 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
ICJNP098xGjgkU	Walkey et al - 84 Gold NPs	84 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.

All Datasets:

Name	Type	Description
1HYbrz5G4pU	test111	test111

Figure 19 Train: Dataset selection

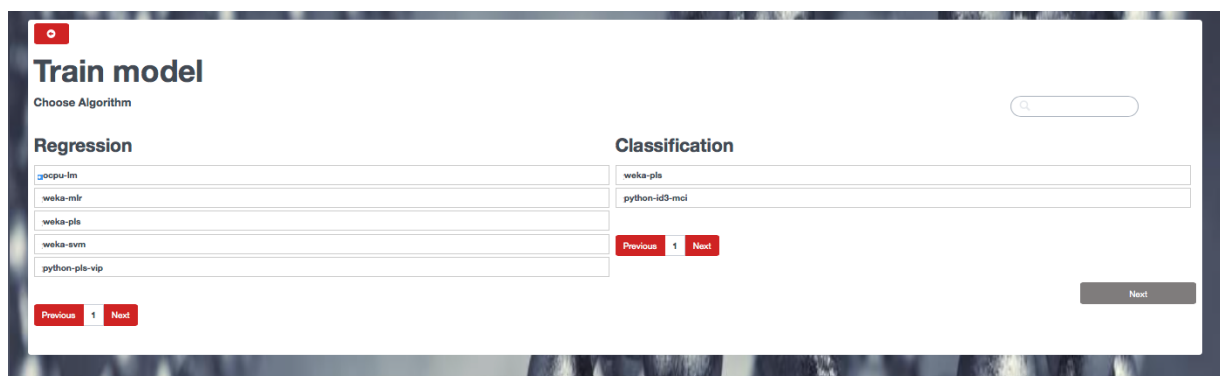
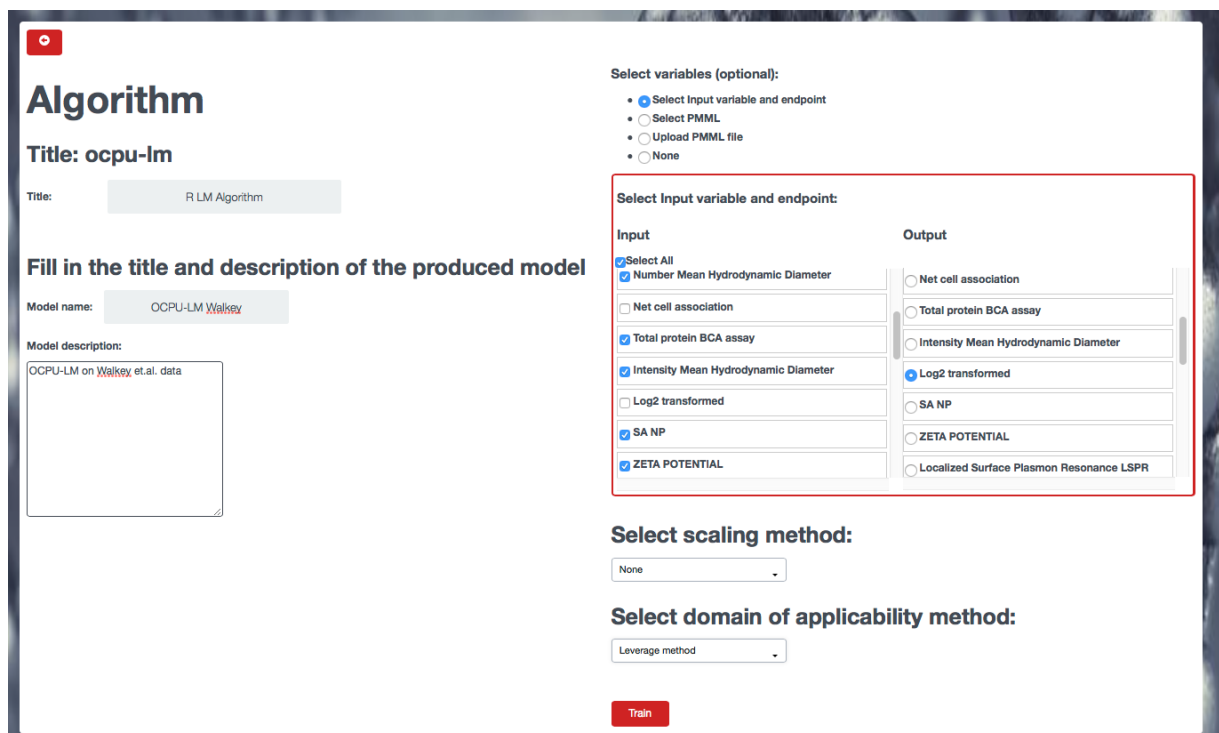


Figure 20 Train: Algorithm selection

Users should access the start page for registered users and select the “**Train a model**” option in **Actions**. This leads to the screen of Figure 19, where the **Walkey et al - 56 Gold NPs** dataset with description: “56 Gold NPs with 25 PhysChem descriptors” should be selected. The screen of Algorithm selection follows (Figure 20). Here **OCPU-lm** should be selected. In Figure 21 the appropriate parameters are given, as shown in the screen (all settings are kept to default). Please note that “Log2 transformed” property has been selected as the end-point, so it need to be excluded from the input variables, as well as the “Net cell association” property, as shown in the figure. Here *Scaling* was not selected and *Domain of Applicability* (DoA) calculations using the *Leverage method* were requested (DoA values closer to 1 show good applicability of model).



Algorithm

Title: **ocpu-lm**

Title:

Fill in the title and description of the produced model

Model name:

Model description:

Select variables (optional):

- ☒ Select Input variable and endpoint
- ☐ Select PMML
- ☐ Upload PMML file
- ☐ None

Select Input variable and endpoint:

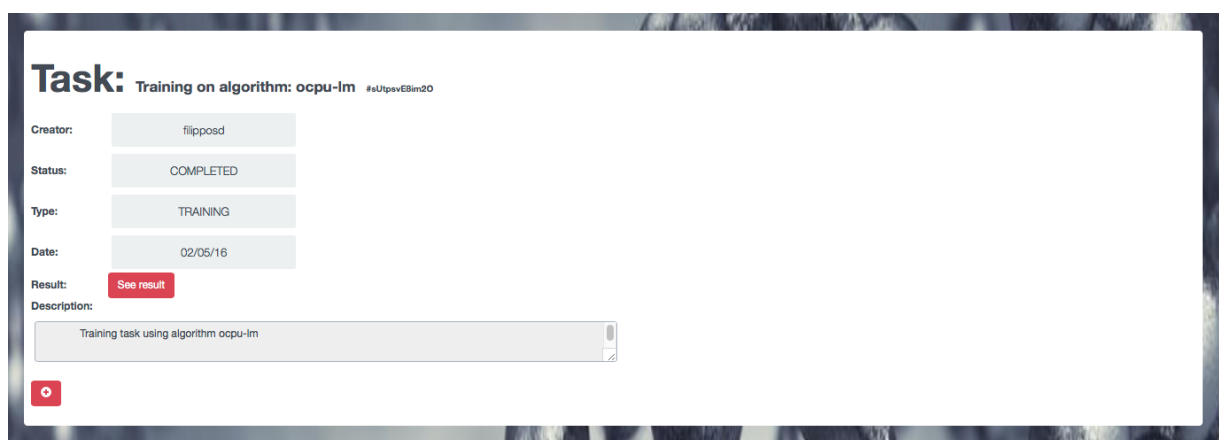
Input	Output
☒ Select All	☐ Net cell association
☒ Number Mean Hydrodynamic Diameter	☐ Total protein BCA assay
☐ Net cell association	☐ Intensity Mean Hydrodynamic Diameter
☒ Total protein BCA assay	☒ Log2 transformed
☒ Intensity Mean Hydrodynamic Diameter	☐ SA NP
☐ Log2 transformed	☐ ZETA POTENTIAL
☒ SA NP	☐ Localized Surface Plasmon Resonance LSPR
☒ ZETA POTENTIAL	

Select scaling method:

Select domain of applicability method:

Figure 21 Train: Algorithm parameters – selection of input-output variables

After an intermediate Task page (Figure 22) the result of the Train action is given, which is the page of the model (Figure 23), which contains full information on the model.



Task: Training on algorithm: **ocpu-lm** #sUtpovE8im20

Creator:

Status:

Type:

Date:

Result:

Description:

Figure 22 Train: Training on algorithm task page

Model:

Name:

Title:

Algorithm:

Description:

Required Features

Dependent Features

Independent Features

Predicted Features

Validations:

Figure 23 Train: Model page with details

A very important aspect here is the generation of the PMML representation of the model, which allows the seamless integration of models into various platforms. This model's PMML file is accessible by users who have logged in (at least as guests) at http://app.jaqpot.org:8000/m_detail?name=LXjNxxY3I3J1.

Predictions will be made by clicking the Predict button and selecting the **Walkey et al - 28 Gold NPs** dataset with description: "28 Gold NPs with 25 PhysChem descriptors". The dataset with predictions is given next (Figure 25).

Select dataset for prediction:

Example datasets:

Name	Title	Description
☐ E20vuMNNTHsp	Gajewicz et al - 10 Metal Oxide NPs	10 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
☐ KIL42dC8JSmp	Walkey et al - 56 Gold NPs	56 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.
☐ class-dummy	Dummy Classification Dataset	This dataset contains classification data
☐ PWY25q3vd5O	Walkey et al - 56 Gold NPs	56 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
☐ UhyeSLaF345pKI	Walkey et al - 84 Gold NPs	84 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
☐ kE0RiswkaCrg	Walkey et al - 28 Gold NPs	28 Gold NPs with 76 protein corona descriptors, used for predicting cellular interaction.
☐ JCUNPD99xGjgkU	Walkey et al - 84 Gold NPs	84 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.
☐ vNxjUZMyv33ITM	Gajewicz et al - 18 Metal Oxide NPs	18 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
☐ xBR5AMG1rOBc	Gajewicz et al - 8 Metal Oxide NPs	8 MeOx NPs with 29 descriptors, used for predicting HaCaT toxicity.
☒ yzsAXE5rLPzz	Walkey et al - 28 Gold NPs	28 Gold NPs with 25 PhysChem descriptors, used for predicting cellular interaction.

Figure 24 Predict: selection of a dataset

Figure 25 Predict: Predictions by model

Similarly to Case 1, the **Walkey et al - 56 Gold NPs** dataset with description: “56 Gold NPs with 76 protein corona descriptors” should be selected, using the **python-pls-vip** algorithm (Figure 26) with the parameters shown in Figure 27, which leads to the model shown in http://app.jagpot.org:8000/m_detail?name=jJPN5UrWvTu6.

Figure 26 Train: Algorithm selection

Algorithm

Title: python-pls-vip

Title: PLS - with VIP scores

Parameters:

latentVariables: 10

Min value:1 Max value:Columns - 1

Fill in the title and description of the produced model

Model name: Walkey PhysChem with PythonPLSV

Model description:

Walkey PhysChem with PythonPLS-VIP

Select variables (optional):

- ☒ Select Input variable and endpoint
- ☐ Select PMML
- ☐ Upload PMML file
- ☐ None

Select Input variable and endpoint:

Input

☒ Select All

☒ P04003

☒ P0C0L5

☒ P05452

☐ Log2 transformed

☒ P01857

☒ P02748

☒ P20R51

Output

☐ P01011

☐ P04003

☐ P0C0L5

☐ P05452

☒ Log2 transformed

☐ P01857

☐ P02748

Select scaling method:

None

Select domain of applicability method:

Leverage method

Train

Figure 27 Train: Algorithm parameters – selection of input-output variables

Predicted values of dataset

#XOkAqU0i0qAa8

Compounds	Log2 transformed	Leverage DoA
G15.Ala-SH_1	-4.82513982144	0.594464451498
G15.CALNN_1	-7.36282507971	0.0
G15.DDT@BDHDA_1	-5.521253232	0.0
G15.DDT@ODA_1	-8.56924967372	0.0
G15.DTNB_1	-5.83875922793	0.0
G15.HDA_1	-3.37489234186	0.0
G15.MBA_1	-3.20512315132	0.0306681058329
G15.MHA_1	-4.63869441755	0.186469033905
G15.MSA_1	-1.45578486172	0.0
G15.NT@DCA_1	-7.27359211576	0.0
G15.NT@PSMA-EDA_1	-6.9812561079	0.0
G15.PAH-SH_1	-2.82328709522	0.0
G15.Phe-SH_1	-7.14910785367	0.366858318975
G15.PVP_1	-7.41536628408	0.0
G15.SPP_1	-5.55120229896	0.0
G15.TP_1	-4.03731742984	0.0
G30.AUT_1	-1.18011822818	0.0
G30.DDT@BDHDA_1	-4.94489214114	0.0
G30.DDT@HDA_1	-4.35167058597	0.0
G30.Met-SH_1	-4.53204856868	0.121814378659

Figure 28 Predict: Predictions by model

Case 3

Use of PMML for determination of input variables and transformations

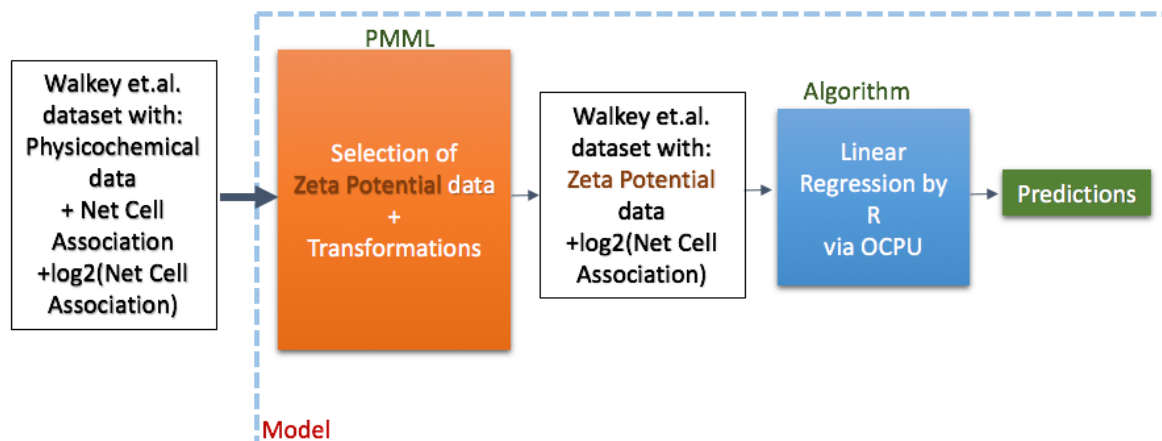
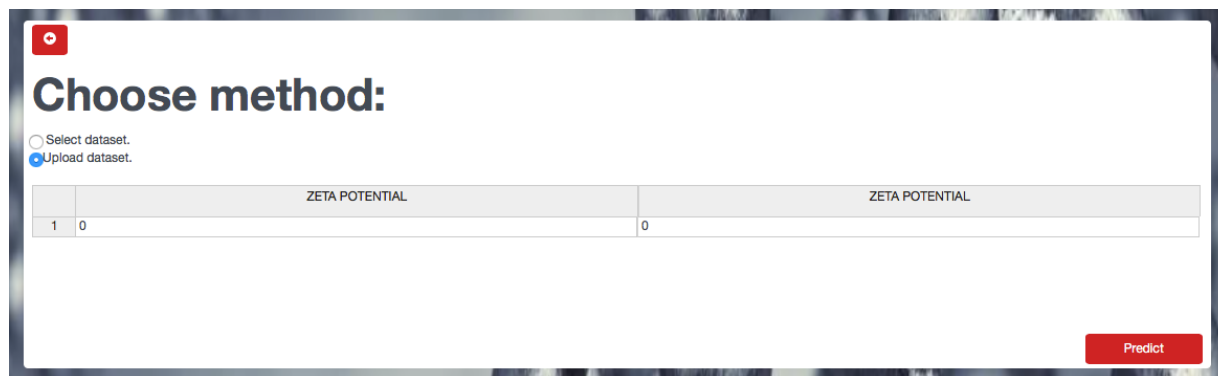


Figure 29 Input variables determination via UI and PMML transformations on the Case 1 dataset

In this case 3, users should use the PMML file in the Workshop webpage in order to select a subset of properties, as shown in the above figure. The input dataset is **Walkey et al - 56 Gold NPs** with description: “56 Gold NPs with 25 PhysChem descriptors” and **Walkey et al - 28 Gold NPs** with description: “28 Gold NPs with 25 PhysChem descriptors” should be used for predictions.

Additionally, another way to get predictions without previously uploading a dataset is by using the “Upload dataset” option shown in Figure 30.



The screenshot shows a web interface titled 'Choose method:'. There are two radio buttons: 'Select dataset.' (unselected) and 'Upload dataset.' (selected). Below the buttons is a table with two columns, both labeled 'ZETA POTENTIAL'. The first row of the table contains the values '1' and '0'. At the bottom right of the interface is a red button labeled 'Predict'.

ZETA POTENTIAL		ZETA POTENTIAL	
1	0	0	

Figure 30 Field to upload a new dataset fore predictions by pasting it using the interface

REFERENCES

1. NTUA eNanoMapper Web services, available at <http://app.jagpot.org:8000>, described at <http://app.jagpot.org:8080/jagpot/swagger/>
2. Hardy, B. et al, Collaborative Development of Predictive Toxicology Applications, Journal of Cheminformatics 2010, 2:7.
3. Walkey, C. D., Olsen, J. B., Song, F., Liu, R., Guo, H., Olsen, D. W. H., Chan, W. C. W. (2014). Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. ACS Nano, 8:3.