DELIVERABLE REPORT D4.5

Design of Experiments and inter-laboratory testing facilities

GRANT AGREEMENT:	604134
ACRONYM:	eNanoMapper
NAME:	eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment
PROJECT COORDINATOR:	Douglas Connect GmbH
START DATE OF PROJECT; DURATION:	1 February 2014; 36 months
PARTNER(s) RESPONSIBLE FOR THIS DELIVERABLE:	NTUA
DATE:	31.10.2016
VERSION:	V1.5

Call identifier	FP7-NMP-2013-SMALL-7
Document Type	Deliverable Report
WP/Task	*WP4/D4.5*
Document ID	eNanoMapper D4.5- Design of Experiments and inter-laboratory testing facilities
Status	Final

Partner Organisations	• Douglas Connect, GmbH (DC) • National Technical University of Athens (NTUA) • In Silico Toxicology (IST) • Ideaconsult (IDEA) • Karolinska Institutet (KI) • European Bioinformatics Institute (EMBL-EBI) • Maastricht University (UM) • Misvik Biology (MB)
Authors	*G. Drakakis, C. Chomenidis, G. Tsiliki, P. Doganis, E. Anagnostopoulou, A. Valsamis, H. Sarimveis (NTUA)* *Reviewed by Barry Hardy (DC).*
Purpose of the Document	*To report on Task 4.5 - Extending statistical and machine learning methods and eNanoMapper modelling functionalities, for the development of experimental design and interlaboratory testing services*
Document History	*1.Table of Contents, 11/07/2016* *2.First draft, 03/08/2016* *3.Second draft 31/08/2016* *4.Third draft 17/10/2016* *5.**Completed version**, 31/10/2016*

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 6

2. INTRODUCTION 7

3. INTERGRATED ANALYSIS API 9

3.1. EXPERIMENTAL DESIGN API 9

ITERATIVE EXPERIMENTAL DESIGN API *9*

FACTORIAL EXPERIMENTAL DESIGN API *13*

3.2. INTERLABORATORY TESTING API 16

4. INTEGRATED ANALYSIS USER APPLICATIONS 19

4.1. EXPERIMENTAL DESIGN USER APPLICATIONS 19

ITERATIVE EXPERIMENTAL DESIGN USER APPLICATION *19*

FACTORIAL EXPERIMENTAL DESIGN USER APPLICATION *21*

4.1. INTERLABORATORY TESTING USER APPLICATIONS 22

5. CONCLUSION 27

6. BIBLIOGRAPHY 28

TABLE OF FIGURES

Figure 1: JQ API found at http://test.jaqpot.org:8080/jaqpot/swagger/
Figure 2: JQ user interface found at http://test.jaqpot.org:8000/
Figure 3: Dataset example for ED in JSON format. The ED algorithm will suggest which of the endpoint 'null' values should be measured.
Figure 4: API for IED experimental design found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/algorithm/trainModel
Figure 5: The new dataset produced for IED experimental design. The column "suggestedTrials" is the output of IED whereas the remaining columns are the original data provided.
Figure 6: API for FED experimental design found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/algorithm/trainModel
Figure 7: FED experimental design output via the API
Figure 8: JSON Example for Interlaboratory testing. Each entry features the same nanoparticle but different owner.
Figure 9: API for interlaboratory testing, found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/interlab/interLabTest
Figure 10: Algorithm options for IED service in JQ User Interface
Figure 11: IED output showing the suggested ENMs for which experiments should be performed (in red colour)
Figure 12: FED algorithm user interface and options
Figure 13: Part of FED algorithm output with suggested experiments (in red colour)
Figure 14: Detailed statistics table (8 tests) from interlaboratory testing report
Figure 15: Actions and warnings raised for particular labs in interlaboratory testing report.
Figure 16: Concealed and real lab names dictionary in inter lab testing report
Figure 17: Histogram of bias estimation for bins of laboratories (auto-binning using Scott's method)
Figure 18: Normal probability plot of expanded uncertainties based on ranks
Figure 19: Normal probability plot of result values. Outliers outside Z-score threshold are annotated with their Z-score value.
Figure 20: Plot of standard deviations against averages with significance values

GLOSSARY

Abbreviation / acronym	Description
A-Optimal	*Average-Optimal*
D-Optimal	*Determinant-Optimal*
I-Optimal	*Integrated-Optimal*
API	*Application Programming Interface*
EC50	*Drug's half maximal Effective Concentration*
ED	*Experimental Design*
eNM	*eNanoMapper*
ENM	*Engineered NanoMaterials*
FED	*Factorial Experimental Design*
IED	*Iterative Experimental Design*
ILT	*Inter-Laboratory Testing*
JQ	*JaqPot Quattro*
JSON	*JavaScript Object Notation*
nQSAR	*nano-QSAR*
OECD	*Organisation for Economic Co-operation and Development*
PCA	*Principal Component Analysis*
PLS	*Partial Least Squares*
PMML	Predictive Model Markup Language
QSAR	*Quantitative Structure-Activity Relationship*
UI	*User Interface*

A list of terms used within the project is accessible at:
http://www.enanomapper.net/library/enm-dictionary

1. EXECUTIVE SUMMARY

This deliverable report describes extensions of the tools available in eNanoMapper's Jaqpot Quattro (JQ) modelling infrastructure, by integrating capabilities for performing optimal experimental design and interlaboratory proficiency testing. In the first case, researchers may optimise the parameters of their experiments either from the very beginning (when they start designing the experiments) or improve their existing workflow in an iterative fashion. In the second case, a coordinator may collect measurements from laboratories to assess potential bias in their pipeline by comparing the values to those of other labs and issue suggestions for improving their experimental procedure. These tools have been developed and deployed as web services built in R and Python respectively. We first indicate and demonstrate these additions via the JQ API and then present a complete walkthrough using the JQ user interface (UI).

2. INTRODUCTION

Experimental design is of crucial importance in the field of ENMs due to the high cost and complexity of conducting such experiments. The significant variability of reported biological and toxicity outcomes on nominally identical materials has challenged the reliability of current testing approaches. A major source of variability is due to the fact that testing methods, experimental conditions, and protocols used by different measurement laboratories are not consistent and standardised. Assessment of the biological and environmental impacts of ENMs can be accelerated, if sources of reliability and irreproducibility can be identified and suitable standard reference materials, assays, and procedures can be established internationally.

Task 4.5 of the eNanoMapper project has developed computational and software tools for supporting, serving and validating inter-laboratory experimental work, interdisciplinary collaboration between experimental and computational scientists and the mutual need for creating quality-driven guidelines for experimental design:

i) The eNanoMapper modelling infrastructure developed in Tasks 4.3, 4.4 and demonstrated in Deliverables 4.3, 4.4 has been extended and tailored to provide directives on which ENMs and under which structures and conditions testing should be carried out to obtain a better exploration of the domain using the minimum possible number of experiments. Optimal experimental design has been advanced beyond today's state of the art by integrating established experimental design methods (such as D-optimal design) with predictive modelling algorithms to develop sequential procedures, taking into account the available experimental data and the model predictions in each iteration.

ii) New functionalities have been developed and offered as ready to use web applications for organising and analysing inter-laboratory tests that allow the assessment of bias within labs by ensemble comparisons of endpoint measurements on the same ENM by the different laboratories. Computational and visualization tools (e.g. z-score and zeta-score histograms) have been developed and provided through the framework for validating the reliability and reproducibility of measurement methods and assessing the performance of individual laboratories, according to the ISO13528.

The new functionalities have been implemented as extensions of both the Jaqpot Quattro (JQ) API (Figure 1, http://app.jaqpot.org:8080/jaqpot/swagger) and the JQ user application (Figure 2, http://test.jaqpot.org:8000/), and they are compatible with the eNanoMapper database and the rest of the modelling services. Section 3 of this report describes the experimental design (ED) and interlaboratory testing (ILT) schemes made available by JQ functionality. Particularly, two ED schemes are implemented, namely the factorial experimental design (FED) which can be used when an experimental group starts planning the experiments and the iterative experimental design (IED) approach which assumes that some experimental data are already available and the group designs the next round of experiments. The ILT scheme is assessing bias among many laboratories based on consensus opinion for the same measurement using 8 statistical tests according to OECD guidelines. A full demonstration of the JQ workflow is given in section 4, with the help of the JQ user interface for the visualization of the functionality of the web services.

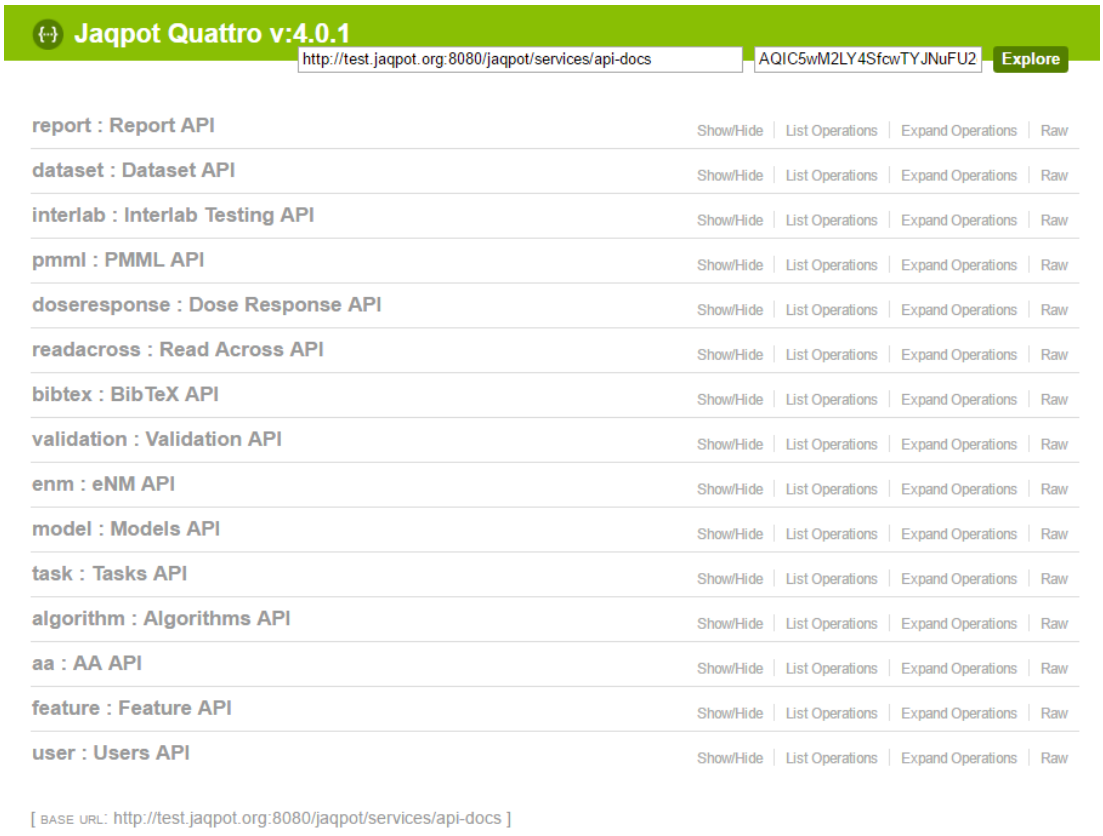


Figure 1: JQ API found at http://test.jaqpot.org:8080/jaqpot/swagger/

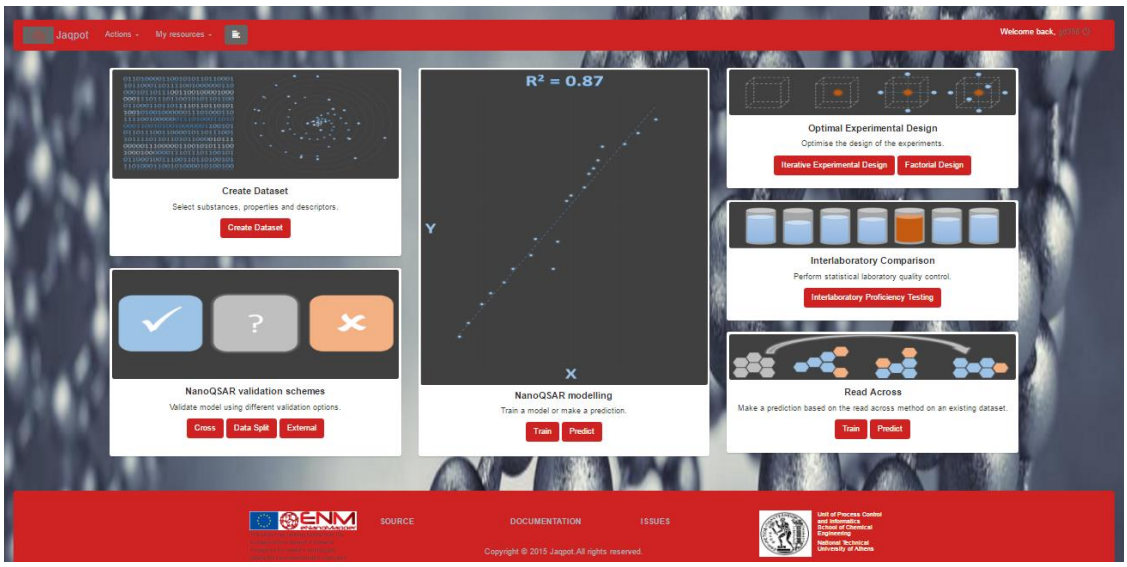


Figure 2: JQ user interface found at http://test.jaqpot.org:8000/

3. INTEGRATED ANALYSIS API

This section describes JQ experimental design and interlaboratory testing algorithm services. As mentioned in deliverable D4.3, these require input data in a standardized format in order to generate their output, and raw experimental data cannot be used directly for such purposes. Therefore, for correct formatting of datasets please refer to the Conjoiner service section in D4.3. Once appropriate datasets exist, users may use the services described below in this document, both on API (described in this section) and UI (see section 4) levels.

3.1. EXPERIMENTAL DESIGN API

As originally introduced in section 5.4 of D4.1, an initial approach of ED algorithm is the FED which builds upon the description of a specific experiment and its variables. This is an ED special case which should be applied when the experimental laboratory starts designing the experiments. In addition, an alternative approach is considered, IED, when some experimental data have been collected and the aim is to suggest the next experiments or 'trials' that need to be conducted. The goal in all cases is to estimate the 'optimal' ED, i.e. the one that would increase performance and would cover the parameter space in the best possible way (Atkinson et al., 2007). The suggested experiments are selected based on three available optimization criteria, namely Determinant-Optimal (D-Optimal), Average-Optimal (A-Optimal), and Integrated-Optimal (I-Optimal). Optimal designs are optimizations based on a chosen optimality criterion and the model that will be fit (section 5.4 of D4.1). For instance, the optimality criterion of D-Optimal design results in minimizing the generalized variance of the parameter estimates for a pre-specified model. Then the optimal set of design runs is chosen from a candidate set of possible design treatment runs. This candidate set of treatment runs usually consists of all possible combinations of various factor levels that one wishes to use in the experiment. The computer algorithm generally uses a stepping and exchanging process to select the set of treatment runs. Federov's heuristic approach was used to optimize the selection speed (Federov, 1972). The eNM ED web service is built as an R openCPU service calling an in-house R package called ExpDesignPkg3.0, which appropriately utilizes the AlgDesign cran package (Wheeler, 2004). The full R code can be found in https://github.com/enanomapper/ExpDesign .

ITERATIVE EXPERIMENTAL DESIGN API

An appropriate dataset in iterative experimental design is either a dataset of untested ENMs, i.e. ENMs for which experimental measurements are available only for input variables (independent variables, such as physicochemical descriptors) or a dataset containing both untested ENMs and tested ENMs, i.e. for some ENMs in the dataset endpoint values have already been measured (assuming in this case that that iterative design has started in a previous iteration). In either case the researcher is seeking to find out the best set of untested ENMs in the dataset, whose endpoint values need to be measured. For example, in the first case, the endpoint variable could be the half maximal effective concentration of a drug (EC50) for 15 untested ENMs and the question would be which experimental EC50 values to measure, whereas in the latter case five, say, EC50 experiments would have been conducted and the question would be to identify the next experimental values to be measured from the remaining ten ENMs. An example screenshot of a dataset suitable for ED in JSON format is shown in Figure 3. This can

be retrieved using GET on a dataset with the "corona-exp" ID using the API, or at http://test.jaqpot.org:8000/data_detail?name=corona-exp via the UI.


```
"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOX
ICITY_SECTION/Log2
+transformed/94D664CFE4929A0F400A5AD8CA733B52E049A688/3ed642f9
-1b42-387a-9966-dea5b91e5f8a" : -5.184,

"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOX
ICITY_SECTION/Net+cell+association/8058CA554E48268ECBA8C98A553
56854F413673B/3ed642f9-1b42-387a-9966-dea5b91e5f8a" : null,

"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOX
ICITY_SECTION/Protein+density/D93618A5D0D5EF5E07CE666BC44D054F
777AE617/3ed642f9-1b42-387a-9966-dea5b91e5f8a" : 0.0,

"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOX
ICITY_SECTION/Total+protein++BCA+assay+/A9C67096CC7A57527BCF9A
A40FD03C39E71647DA/3ed642f9-1b42-387a-9966-dea5b91e5f8a" :
2.927,

"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOX
ICITY_SECTION/Total+surface+area++SAtot+/52D93BC3B68F26C8E787C
C7A05E5130A23164405/3ed642f9-1b42-387a-9966-dea5b91e5f8a" :
11.0
    }
  }, {
    "compound" : {
      "name" : "G15.AHT_5",
      "URI" :
"https://apps.ideaconsult.net/enmtest/substance/FCSV-0e1a05ec-
6045-3419-89e5-6e48e1c62e3c"
    },
```


Figure 3: Dataset example for ED in JSON format. The ED algorithm will suggest which of the endpoint 'null' values should be measured.

The IED scheme as implemented for the eNM platform is based on the work published by Brandmaier et al. (2012) and Brandmaier and Tetko (2013), i.e. it is an adaptive approach that combines optimal criteria (D-Optimal, A-Optimal, I-Optimal) with partial least squares (PLS) and leave-one-out cross validation or principal component analysis (PCA). Particularly, the realization of the nano-chemical space is built with either PLS components (when tested ENMs are included in the dataset) or PCA components (when the dataset contains only untested ENMs). The latent variables from PLS are ranked by their correlation to the endpoint variable, whereas the principal components from PCA are ranked by their variance in the descriptor space (Brandmaier et al., 2013).

The implemented algorithms can be retrieved via the API by using GET algorithm by ID and type the following ids "ocpu-expdesign2-xy" or "ocpu-expdesign2-x", depending on whether endpoint experimental measurements are or are not included in the data. In order to use this service, users must POST several parameters under the algorithm POST by ID option (Figure 4). "Title" and "description" are optional, but users are urged to fill them in as rigorously as possible for completeness. The "dataset URI" must be experimental design-compliant, meaning that the endpoint variable is optional but if included they should contain missing "null" values (i.e. some empty cells in the endpoint column of the data matrix from which IED should select the optimal set). The dataset "corona-exp" mentioned above is a

representative example. The service automatically detects whether an endpoint variable “Prediction feature” with missing values is available.

The IED service awaits five parameter values; the list is given below. Please note, however, that for completeness we include in parentheses the terms as they appear in JQ UI and will be further discussed in section 4.1.:

1. “nTrials” (“Number of Trials”), a numeric value indicating number of trials suggested. If not supplied then the default value 11 is used, however in the current version a value greater than 4 is expected.
2. “criterion” (“Optimal design criterion”) a character value to indicate which optimal design to apply (possible values are “D”, “A”, “I”).
3. “r2.threshold” (“R2 threshold”) a numeric value indicating the R2 threshold value. If the data supplied provides R2 value greater than the threshold value, a warning message is returned.
4. “form” (“Design formula”), a string indicating the formula of the design as defined in AlgDesign R package (possible formulas are “linear”, “quad” for quadratic models, “cubic”, “cubicS” for a non-singular representation of a cubic model)
5. “newY” (“New Dependent Variable”), a string name of the new dependent feature, when the dataset does not contain any variable with missing values.

Parameters should be given in JSON format, for example:

{"nTrials":[5],"criterion":["D"],"r2.threshold":[0.8],"form":["linear"],"newY":"newY"}

Please note that the last variable is only needed for the ocpu-expdesign2-x service, and also that the current version awaits at least 4 samples in the data. By clicking on “Try it out!” a model is created with IED results and its ID is returned. Then the API POST model by id service is used to create a dataset with the first column being the suggested trials of IED analysis.

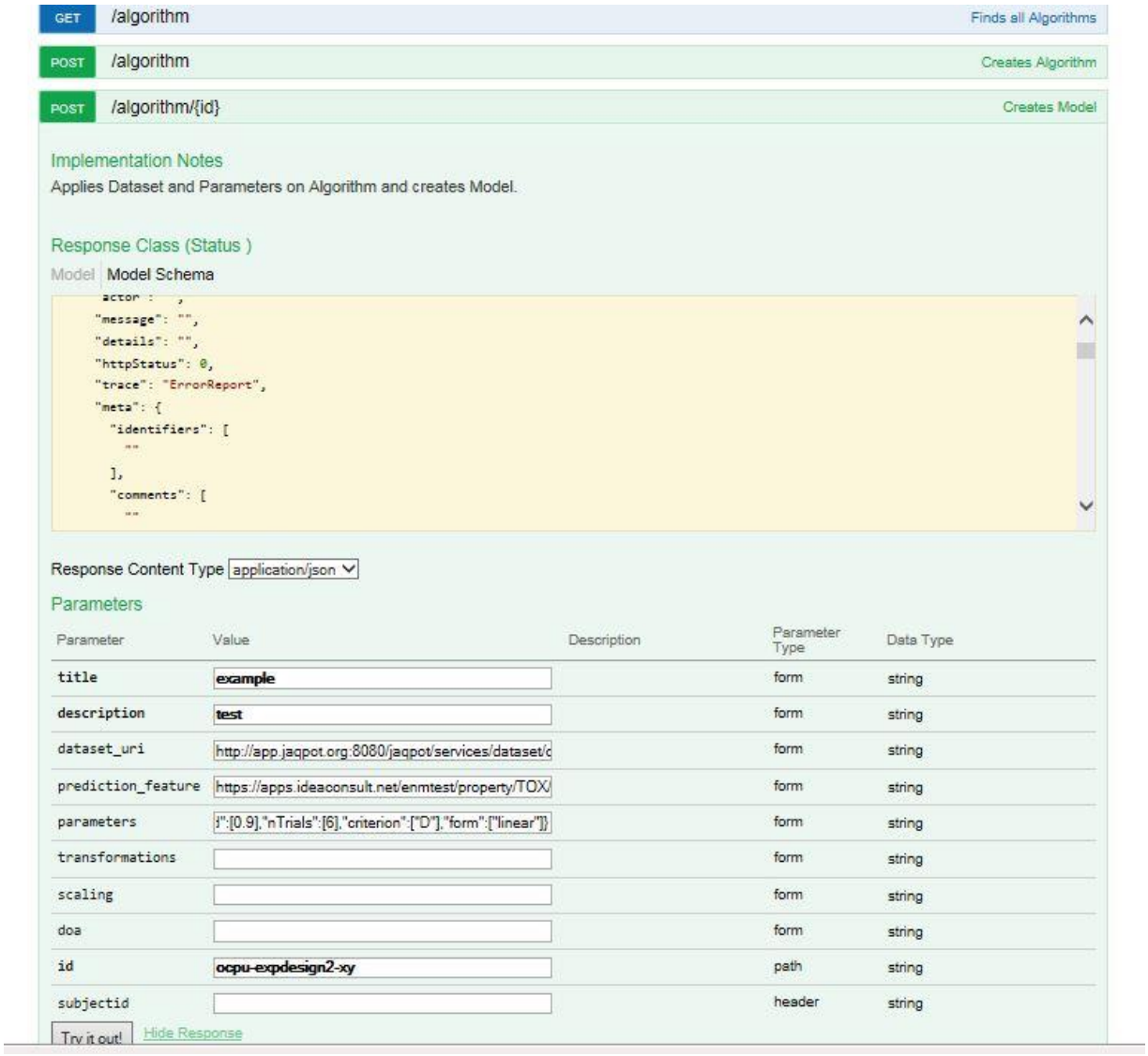
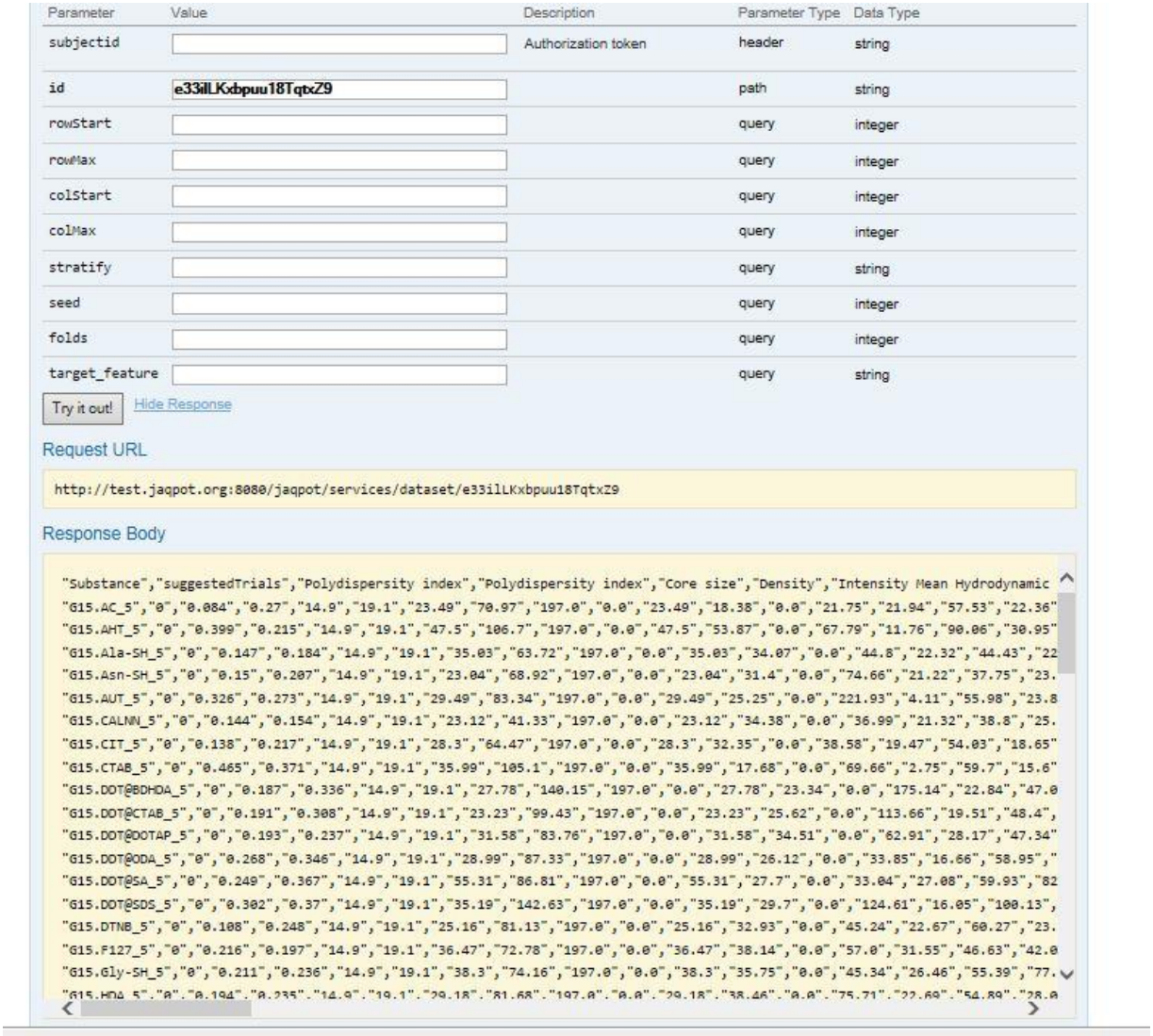


Figure 4: API for IED experimental design found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/algorithm/trainModel

The result is a dataset, which additionally to the dataset supplied, contains the suggested trials (i.e. the untested ENM for which endpoint values should be measured, Figure 5). This object is a binary vector with 1 indicating suggested trial. Additional information is also included: the R^2 value of the model if the endpoint variable is given, the confounding effect (i.e. the diagonality of the design, excluding the constant, if any), and the normalized variance of the design.



Parameter	Value	Description	Parameter Type	Data Type
subjectid		Authorization token	header	string
id	e33ilLKxbpuu18TqtxZ9		path	string
rowStart			query	integer
rowMax			query	integer
colStart			query	integer
colMax			query	integer
stratify			query	string
seed			query	integer
folds			query	integer
target_feature			query	string

Try it out! Hide Response

Request URL

http://test.jaqpot.org:8080/jaqpot/services/dataset/e33ilLKxbpuu18TqtxZ9

Response Body

"Substance","suggestedTrials","Polydispersity index","Polydispersity index","Core size","Density","Intensity Mean Hydrodynamic
"G15.AC_5","0","0.084","0.27","14.9","19.1","23.49","70.97","197.0","0.0","23.49","18.38","0.0","21.75","21.94","57.53","22.36"
"G15.AHT_5","0","0.399","0.215","14.9","19.1","47.5","106.7","197.0","0.0","47.5","53.87","0.0","67.79","11.76","90.06","30.95"
"G15.Ala-SH_5","0","0.147","0.184","14.9","19.1","35.03","63.72","197.0","0.0","35.03","34.07","0.0","44.8","22.32","44.43","22
"G15.Asn-SH_5","0","0.15","0.207","14.9","19.1","23.04","68.92","197.0","0.0","23.04","31.4","0.0","74.66","21.22","37.75","23.
"G15.AUT_5","0","0.326","0.273","14.9","19.1","29.49","83.34","197.0","0.0","29.49","25.25","0.0","221.93","4.11","55.98","23.8
"G15.CALNN_5","0","0.144","0.154","14.9","19.1","23.12","41.33","197.0","0.0","23.12","34.38","0.0","36.99","21.32","38.8","25.
"G15.CIT_5","0","0.138","0.217","14.9","19.1","28.3","64.47","197.0","0.0","28.3","32.35","0.0","38.58","19.47","54.03","18.65"
"G15.CTAB_5","0","0.465","0.371","14.9","19.1","35.99","105.1","197.0","0.0","35.99","17.68","0.0","69.66","2.75","59.7","15.6"
"G15.DOT@BDHDA_5","0","0.187","0.336","14.9","19.1","27.78","140.15","197.0","0.0","27.78","23.34","0.0","175.14","22.84","47.0
"G15.DOT@CTAB_5","0","0.191","0.308","14.9","19.1","23.23","99.43","197.0","0.0","23.23","25.62","0.0","113.66","19.51","48.4",
"G15.DOT@DOTAP_5","0","0.193","0.237","14.9","19.1","31.58","83.76","197.0","0.0","31.58","34.51","0.0","62.91","28.17","47.34"
"G15.DOT@ODA_5","0","0.268","0.346","14.9","19.1","28.99","87.33","197.0","0.0","28.99","26.12","0.0","33.85","16.66","58.95","
"G15.DOT@SA_5","0","0.249","0.367","14.9","19.1","55.31","86.81","197.0","0.0","55.31","27.7","0.0","33.04","27.08","59.93","82
"G15.DOT@SDS_5","0","0.302","0.37","14.9","19.1","35.19","142.63","197.0","0.0","35.19","29.7","0.0","124.61","16.05","100.13",
"G15.DTNB_5","0","0.108","0.248","14.9","19.1","25.16","81.13","197.0","0.0","25.16","32.93","0.0","45.24","22.67","60.27","23.
"G15.F127_5","0","0.216","0.197","14.9","19.1","36.47","72.78","197.0","0.0","36.47","38.14","0.0","57.0","31.55","46.63","42.0
"G15.Gly-SH_5","0","0.211","0.236","14.9","19.1","38.3","74.16","197.0","0.0","38.3","35.75","0.0","45.34","26.46","55.39","77.

Figure 5: The new dataset produced for IED experimental design. The column "suggestedTrials" is the output of IED whereas the remaining columns are the original data provided.

FACTORIAL EXPERIMENTAL DESIGN API

If no more than a description of the experiments is available and researchers have decided on the protocol of the experiments including which variables are of interest and what are the levels that should be tested, a factorial design could be used to construct the design matrix, that is a table with all possible combinations of variables and levels. Then, relatively to the above, the optimal criteria application via IED and "ocpu-expdesign2-x" algorithm should decide on which are the combinations that are most relevant to the question of interest. For instance, if the question is to estimate the most relevant parameters and levels to the EC50 endpoint variable, and the parameters considered where three with three levels each, the FED would construct a 3^3 full factorial design matrix, then PCA will be applied and Federov's exchange algorithm would suggest an optimal design given the selection criterion ("D", "A", "I") specified.

The implemented FED algorithm can be retrieved via the swagger API by using GET algorithm by ID and type “ocpu-expdesign2-noxy”. In order to use this service, users must POST several parameters under the algorithm POST by ID option “title” and “description” are optional, but users are urged to fill them in. Users must also POST the following parameters (terms in parentheses are those appearing in JQ UI):

1. “nVars”, a numeric value indicating the number of variables included in the design,
2. “varNames”, an array of length “nVars” with strings for the variable names included in the design,
3. “levels”, a list of arrays corresponding to particular levels per variable, it can be an integer if levels of all variables are equal.
4. “nTrials” (“Number of trials”), the number of suggested trials. If not specified, the default value 11 is used.
5. “criterion” (“Optimal Design Criterion”), possible options are “D”, “A”, and “I” (see section 3.1.1),
6. “form” (“Design formula”) possible options are “linear”, “quad”, “cubic”, “cubicS”,
7. “factors”, a numeric array which indicates in ascending order which are the factor/categorical variables (if all variables are numeric “0” should be declared). In the current version only numeric variables are accepted.
8. “newY”, a string indicating the name of the dependent variable to be measured

Parameters should be given in JSON format, for example:

{"nVars":[3],"varNames":["a","b","c"],"levels":{"a":[30,40],"b":[10,20,30],"c":[1,2,3]},
"nTrials":[5],"criterion":["D"],"form":["linear"],"factors":[0],"newY":["CA"]}

By clicking on “Try it out!” a model is created with FED results and its ID is returned. Figure 6 shows a screenshot of the POST API FED service. Then an empty dataset should be created by using the POST dataset service. The datasetUri produced and the model id are given to the POST by id model service to produce the full factorial design with an extra column, called “suggestedTrials”, indicating with “1” the new experiments to be performed, and “0” otherwise (Figure 7).

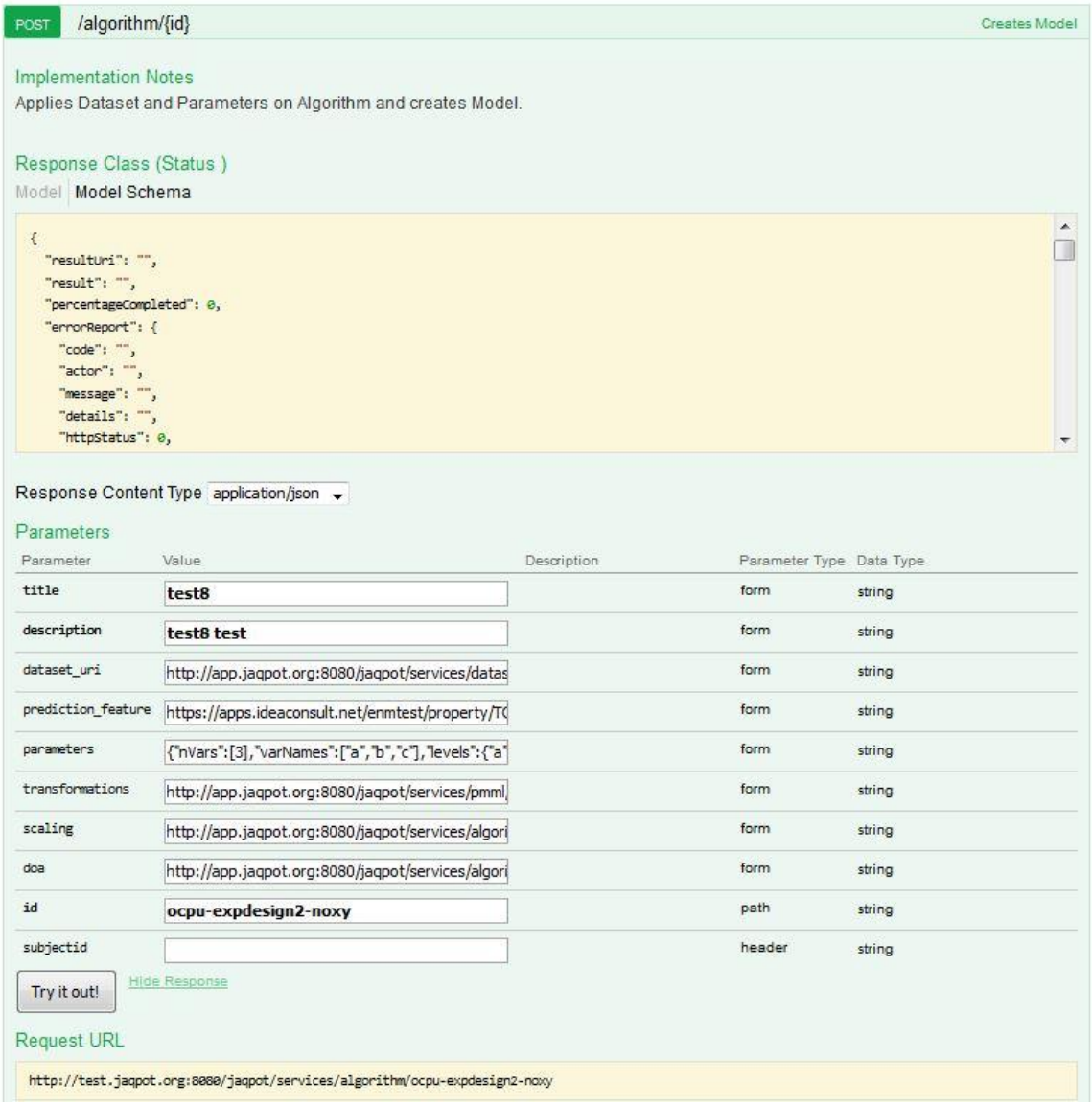


Figure 6: API for FED experimental design found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/algorithm/trainModel

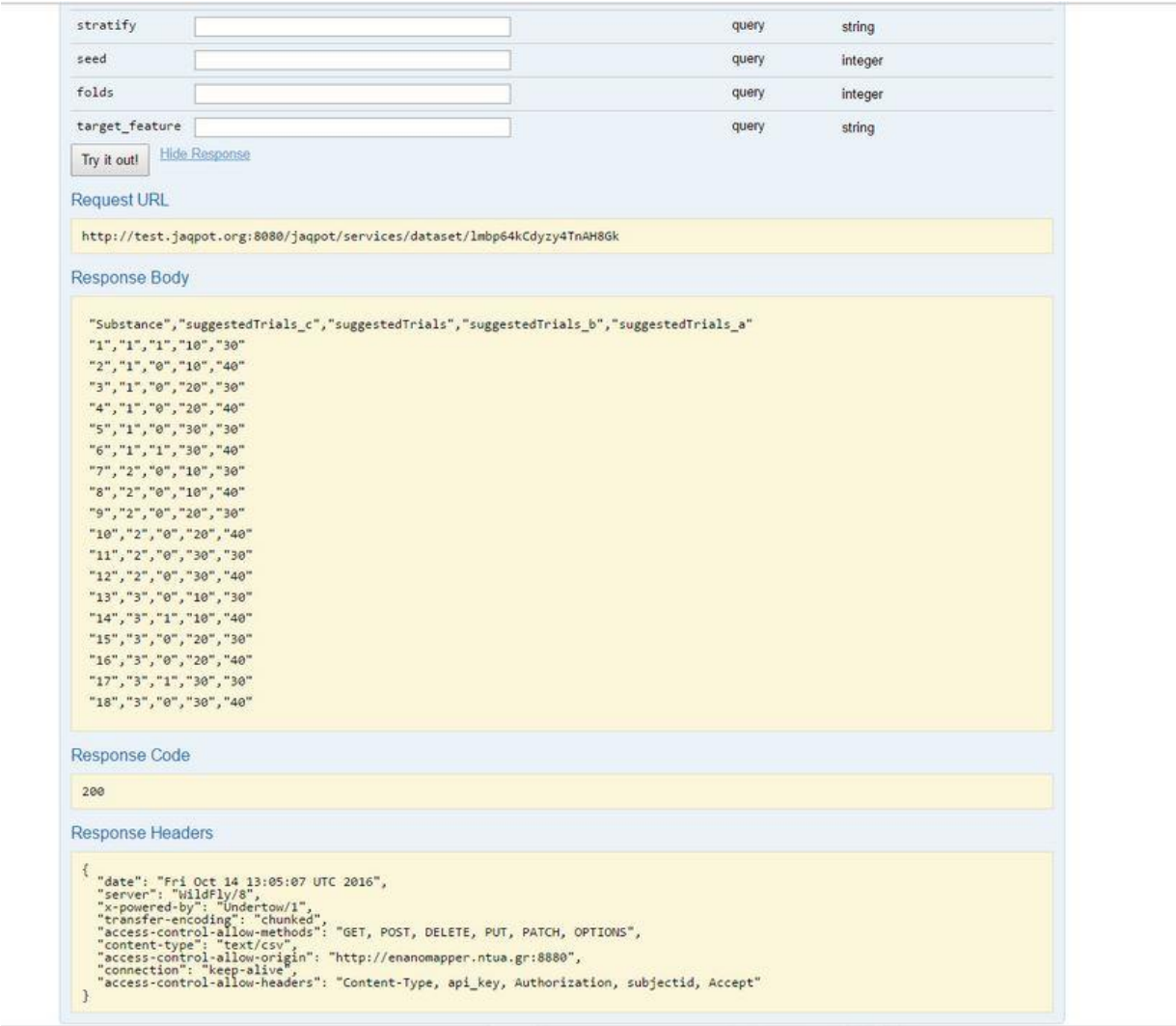


Figure 7: FED experimental design output via the API

3.2. INTERLABORATORY TESTING API

In this section we will describe the usage of the interlaboratory testing functionality of JQ. Interlaboratory proficiency testing is a series of statistical tests used to assess bias in laboratories and increase repeatability of experiments according to international standards (ISO 13528:2015). This can be carried out by processing the reported measurement results of laboratories on either reference materials for which an expected range of values exists, or on the consensus outcomes of a group of laboratories.

A coordinator for the proficiency testing is responsible for collecting the results from each laboratory, in order to provide homogenous datasets and anonymise the laboratories for the report. Based on the JQ system architecture and the functionalities provided, it is evident that in order to carry out such a test a dataset must be created, for which each entry in the dataset should be a different lab for measuring the same endpoint. JQ supports single and multiple entries, as well as uncertainties, should these have been

reported by the participating laboratories. An example screenshot of a dummy dataset suitable for interlaboratory testing in JSON format is shown in Figure 8. It is evident that it contains entries of the same nanoparticle by different owners and allows multiple measurements as well as uncertainties. This can be retrieved using GET dataset by ID ("interlab-dummy") using the API, or at http://test.jaqpot.org:8000/data_detail?name=interlab-dummy via the UI.


```
"dataEntry": [
    {
      "compound": {
        "name": "G15.AC_5",
        "ownerUUID": "FCSV-8b479138-4775-3aba-b9cc-f01cc967d42b",
        "URI":
        "https://apps.ideaconsult.net/enmtest/substance/FCSV-8b479138-4775-3aba-b9cc-f01cc967d42b"
      },
      "values": {
"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed+uncertainties
/94D664CFE4929A0F400A5AD8CA733B52E049A688/3ed642f9-1b42-387a-9966-dea5b91e5f8a": [ 20 ],
"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed/94D664CFE4929
A0F400A5AD8CA733B52E049A688/3ed642f9-1b42-387a-9966-dea5b91e5f8a": [ 180, 175, 178 ]
      }
    },
    {
      "compound": {
        "name": "G15.AC_5",
        "ownerUUID": "FCSV-8b479138-4775-3aba-b9cc-fasdfasdfasdf",
        "URI":
        "https://apps.ideaconsult.net/enmtest/substance/FCSV-0e1a05ec-6045-3419-89e5-6e48e1c62e3c"
      },
      "values": {
"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed+uncertainties
/94D664CFE4929A0F400A5AD8CA733B52E049A688/3ed642f9-1b42-387a-9966-dea5b91e5f8a": [ 0 ],
"https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed/94D664CFE4929
A0F400A5AD8CA733B52E049A688/3ed642f9-1b42-387a-9966-dea5b91e5f8a": [ 203, 203, 203 ]
      }
    },
...
```


Figure 8: JSON Example for Interlaboratory testing. Each entry features the same nanoparticle but different owner.

Based on their performance (8 discrete statistical tests), laboratories may receive action or warning signals, which include checking the measurement procedures, competence of staff and performing calibration of the laboratory equipment. Labs which do not receive any signals are considered to pass the statistical tests.

Depending on the number of measurements provided by each lab, all datasets entered are subjected to ISO-provided algorithms for defining the robust average and robust standard deviation. Then the assigned value and uncertainty are calculated, the latter depending on whether the laboratories have provided their own or not.

Once these values have been calculated the process continues by calculating performance statistics using ISO-defined bins (accepted ranges) and thresholds for estimates of laboratory bias, percentage differences, ranks, z-scores, En scores, z' scores, ζ-scores and Ez-scores. Values outside the bins or above/below these thresholds will receive warning or action signals. Several of these measurements are visualised for the user in the form of histograms of reported scores, bar plots of bias, Probability Plots, etc.

The swagger API for interlaboratory testing can be found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/interlab/interLabTest , as shown in Figure 9. In order to

use this service, users must POST several parameters. 'Title' and 'description' are optional, but users are urged to fill them in as rigorously as possible for completeness. The 'dataset URI' must be interlaboratory testing-compliant, meaning that all measurements should be on the same nanoparticle from different owners. For this tutorial we have created and used "interlab-dummy". By viewing this dataset, it can be seen that the "prediction feature" is "Log2 transformed", whose URI is https://apps.ideaconsult.net/enmtest/property/TOX/UNKNOWN_TOXICITY_SECTION/Log2+transformed/94D664CFE4929A0F400A5AD8CA733B52E049A688/3ed642f9-1b42-387a-9966-dea5b91e5f8a . The service automatically detects whether another variable is present and assigns it as the reported uncertainty, should it exist. By clicking on "Try it out!" a report is created and its ID is returned. This can be viewed under the GET report by ID API, or at http://test.jaqpot.org:8000/report?name=zZxyLhWg28GXHrB with the help of the UI, shown later in this document.

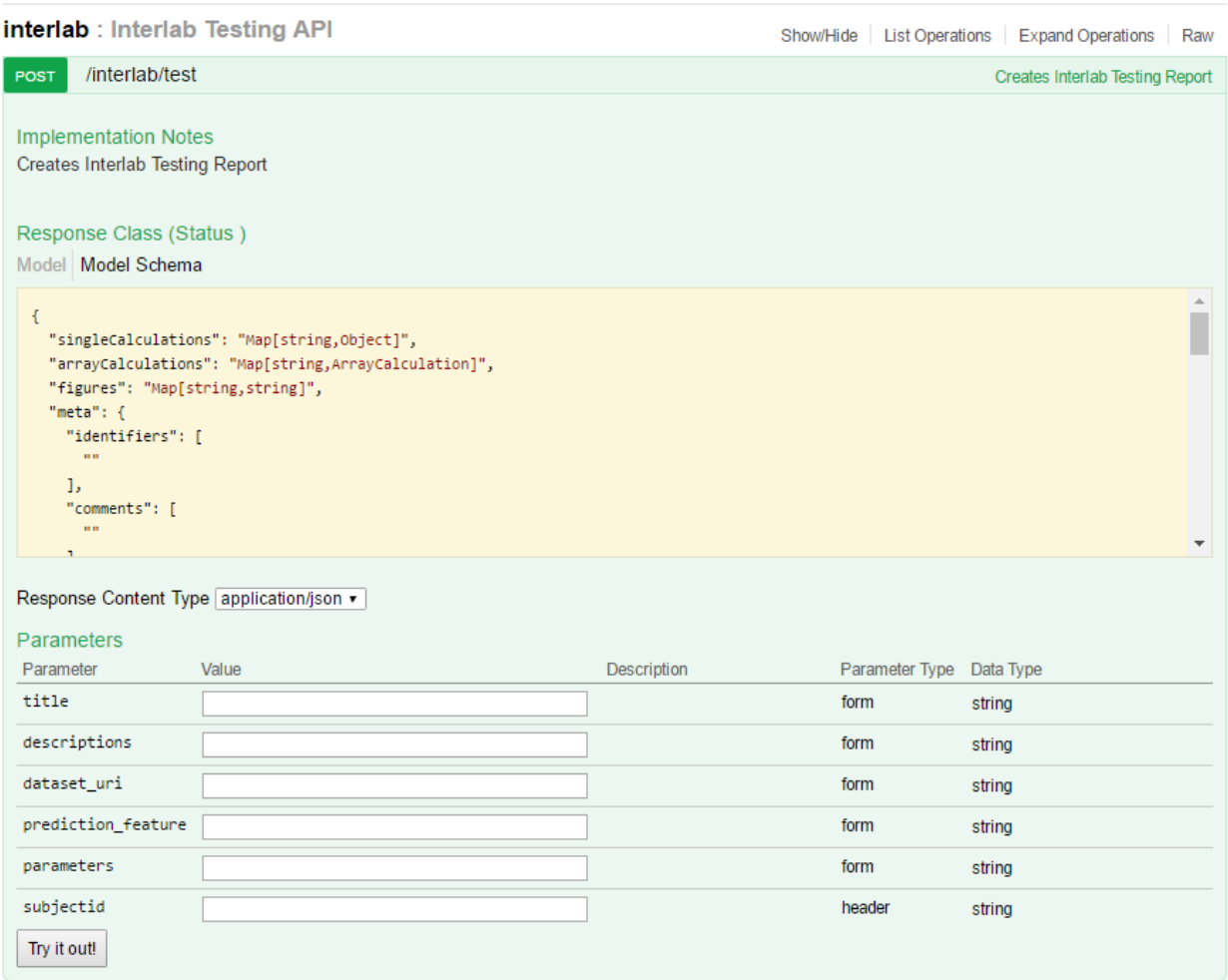


Figure 9: API for interlaboratory testing, found at http://test.jaqpot.org:8080/jaqpot/swagger/#!/interlab/interLabTest

4. INTEGRATED ANALYSIS USER APPLICATIONS

This section of deliverable 4.5 presents the user applications which have been developed based on the APIs presented in Section 3 for optimal experimental design and interlaboratory comparison. The easy-to-use web applications support the iterative and complementary work of experimental and modelling groups and have been integrated in the JQ modelling user interface.

4.1. EXPERIMENTAL DESIGN USER APPLICATIONS

As already explained in the previous section, ED algorithms assist users to either design additional experiments of an ongoing research project by calling IED or initial experiments by calling FED. In the first case given the data already available, a leave-one-out cross validation scheme is employed for a PLS regression model when the dataset contains some tested ENMs or a PCA representation of the data when the dataset does not contain any endpoint measurements. The suggested experiments are selected based on three available optimization criteria, namely D-Optimal, A-Optimal, I-Optimal. In the latter case, a full factorial design is built and the final choice of suggested experiments is based upon the PCA components of the created dataset optimised using one of the above mentioned criteria.

ITERATIVE EXPERIMENTAL DESIGN USER APPLICATION

Users must first specify the dataset that needs to be analysed, which would normally be a data matrix with some, or in the extreme case, all of the endpoint values missing. Subsequently, the user must specify the algorithmic parameters as shown in Fig. 10. Particularly, the user can name the analysis under the “Title” option, specify the number of suggested trials returned by the algorithm “Number of Trials” (that should be a number smaller or equal to the missing endpoint values), the optimization criterion used (“D”, “A”, “I”), a threshold value for the coefficient of determination R^2 of the existing dataset (a checkpoint service to alert the user in case this condition is already met), the formula used to create the design in “Design formula” (“linear”, “quadratic”, “cubic”, or a non-singular representation of a cubic model “cubicS”), and under the “New Dependent Variable” field specify the name of the new feature created to store the suggested trials. Please note that if the R^2 value of the data (non-empty endpoint values) is higher than the threshold R^2 value supplied, the dataset is assumed to be rich in information (no additional experiments are suggested) and the user is prompted to use the dataset for modelling purposes.

Under the “Select variables” field the user selects independent (input) variables, and he/she is able to define transformations using the PMML format (see D4.3). If the user wishes to keep all independent variables, then the option “Select endpoint only” should be chosen. The variable with missing information is automatically detected and is displayed as the endpoint variable. If the dataset does not contain any variable with missing information, the user is prompted to generate and name a new dependent variable. “Select scaling method” field performs data scaling.

Figure 11 shows part of the results obtained for IED experimental design algorithm with the defaults parameters, when applied to a proteomics dataset consisting of 28 gold untested ENMs and 76 proteins (dataset id "#kpNKB3vSZKkYxCxiaZLu"), which are considered as independent variables, while the net cell association to A549 human lung epithelial carcinoma cells is considered as the endpoint of interest. Experimental design statistics are reported in the top of the screen to help the user assess the results. The table underneath includes all 28 ENMs considered in the experimental design and an indication column named 'Prediction feature' (if no other name was supplied in the "New Dependent Variable" parameters field) which shows in red which should be the next experiments. After performing those experiments, the user can directly type results in this form and re-run the service to identify additional experiments.

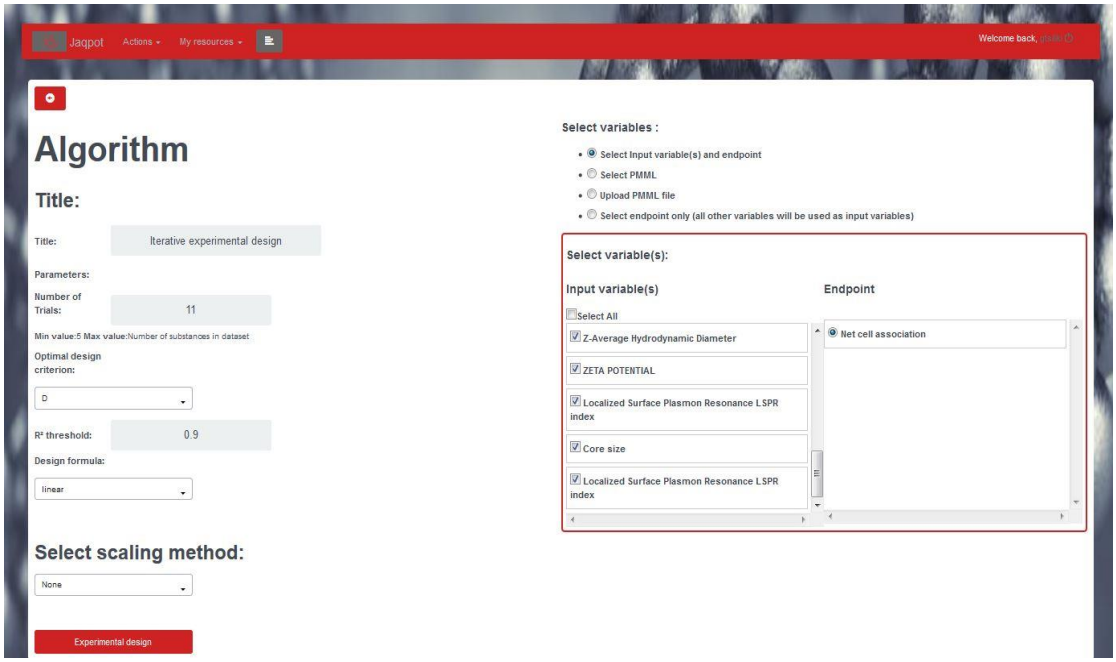


Figure 10: Algorithm options for IED service in JQ User Interface

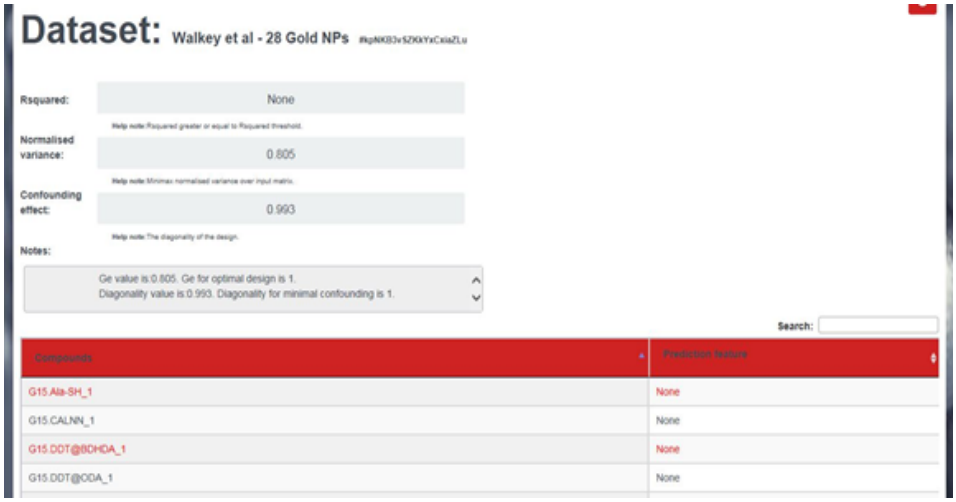


Figure 11: IED output showing the suggested ENMs for which experiments should be performed (in red colour)

FACTORIAL EXPERIMENTAL DESIGN USER APPLICATION

In this case users need to specify all the independent variables, the parameters and characteristics of the experiments they wish to conduct (Fig. 12). The following fields are available: “Number of trials” (the returned number of suggested trials), “Optimal Design Criterion” (the optimization criterion used, one of “D”, “A”, “I”), and “Design formula” (the formula used to create the design; one of “linear”, “quadratic”, “cubic”, “cubicS”). An interactive table is included where the user is expected to define all remaining parameters of his/her experiment. Particularly, the user should specify the names of independent variables under the column “Variables”, the type of each independent variable under the column “Type” (currently only numeric variables are expected), and the explicit levels per variable in the remaining columns. Rows/columns can be adjusted according to the experiment described. Also, a “Dataset name” and “Dataset description” should be filled in, so that the user can easily retrieve the produced dataset in the future.

The output is shown in (Fig. 13). A full factorial design of 4 variables with different numbers of levels is created, the column named “Prediction feature” is the new column of the dataset created indicating the endpoint variables and awaiting the new experiment values. The suggested ENMs for which experiments should be performed are indicated in red. The remaining columns indicate the levels for each of the 4 variables considered in this case. There is the option available to insert values in the “Prediction feature” field and subsequently “Save Dataset” for further analysis, either by IED or any other modelling algorithm.

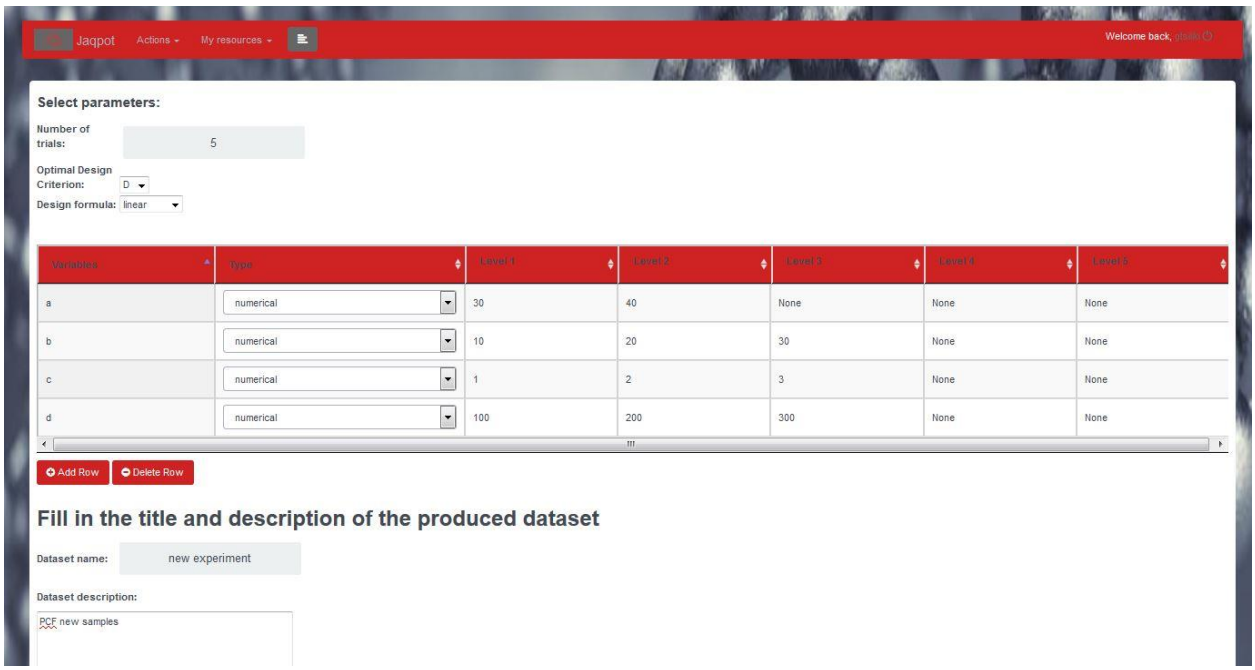
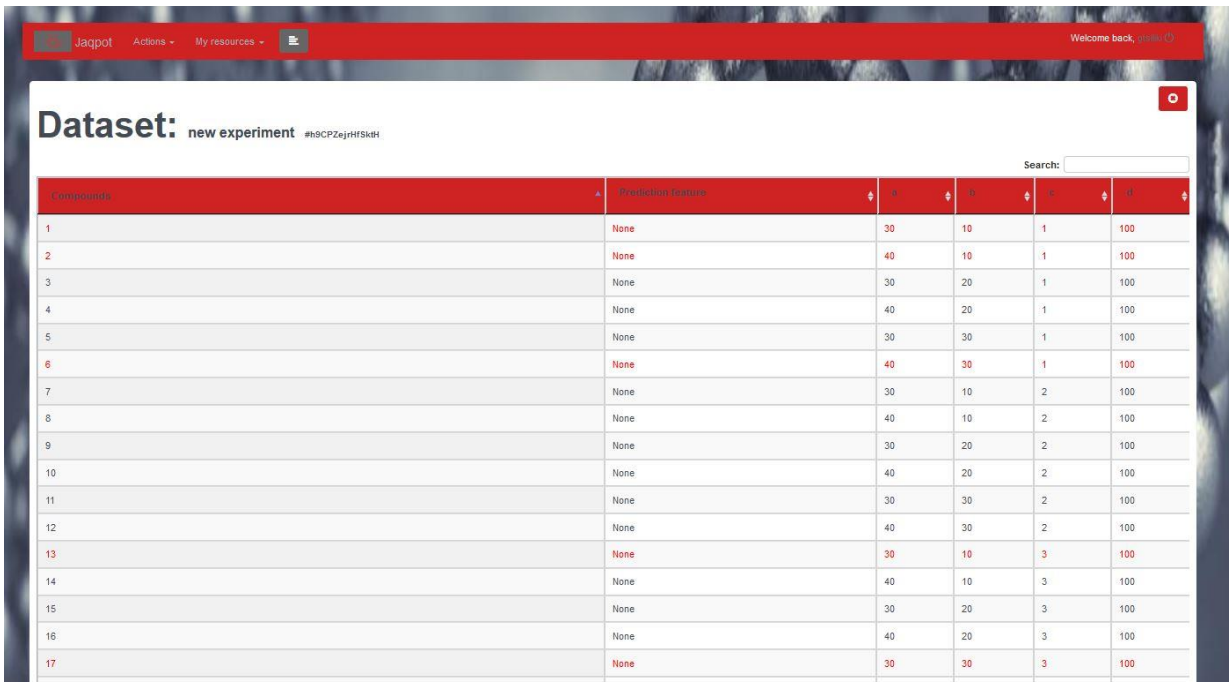


Figure 12: FED algorithm user interface and options



Compounds	Prediction feature	a	b	c	d
1	None	30	10	1	100
2	None	40	10	1	100
3	None	30	20	1	100
4	None	40	20	1	100
5	None	30	30	1	100
6	None	40	30	1	100
7	None	30	10	2	100
8	None	40	10	2	100
9	None	30	20	2	100
10	None	40	20	2	100
11	None	30	30	2	100
12	None	40	30	2	100
13	None	30	10	3	100
14	None	40	10	3	100
15	None	30	20	3	100
16	None	40	20	3	100
17	None	30	30	3	100

Figure 13: Part of FED algorithm output with suggested experiments (in red colour)

4.1. INTERLABORATORY TESTING USER APPLICATIONS

CREATING AN INTERLABORATORY TESTING REPORT

Users may either click on the “Interlaboratory Proficiency Testing” button found on the right hand side of the main JQ UI page, or alternatively go under the “Actions” menu and click the “Interlab Testing” option. A screen will then prompt the user to select a dataset. Once a dataset suitable for interlaboratory testing (same nanoparticle and different laboratory for each entry) has been selected, users are prompted to provide a title and description for the report to be generated. Users (or coordinators in this case) are urged to be as thorough as possible with their description in order to be able to easily locate their report at any time. By clicking on the “Interlab Testing” button a task is created. Once finished, users can view the full report and download it as a pdf file. This contains single calculations, array calculations and figures. Single calculations include the title, description, robust average and standard deviation and calculated uncertainty of the assigned value. Array calculations contain detailed statistics, action and warning signals, ranks, dictionary containing lab real name and dummy name for anonymity in the report.

For this tutorial, an example dataset has been added to the JQ "example datasets" for interlaboratory testing, named "interlab-dummy" with the description "Interlab Testing Dataset". The dataset can be viewed in full under "My resources", or by following the link http://test.jaqpot.org:8000/data_detail?name=interlab-dummy . In this example, 10 labs have reported 3 measurements plus their estimate of uncertainty. After applying the JQ interlaboratory service, the following report is produced: http://test.jaqpot.org:8000/report?name=zZxyLhWg28GXHrB. At the top of the report, users can see the report unique ID, the title, description, robust average, robust standard deviation and calculated uncertainty of the assigned value, as mentioned above. Underneath these entries we can see the array calculations. The most important are the "Detailed Statistics" (shown in Fig. 14), the "Action Signals" and "Warning Signals" for labs which have failed the statistical tests (shown in Fig. 15) and the "Lab Real Names" dictionary containing lab real name and dummy name for anonymity, should the coordinator choose to share the rest of the report with the participating labs (shown in Fig. 16).

Detailed Statistics

	Lab	z-score	En value (95%)	En value (99%)	z' score	zeta score	Ez- score (95%)	Ez+ score (95%)	Ez- score (99%)	Ez+ score (99%)
Row_1	Lab_1	0.08	0.03	0.02	0.07	0.06	0.42	0.42	0.41	0.41
Row_10	Lab_10	-0.44	-0.14	-0.1	-0.39	-0.29	0.22	0.22	0.27	0.27
Row_2	Lab_2	1.81	1.71	1.14	1.6	3.42	2.71	2.71	2.14	2.14
Row_3	Lab_3	-0.38	-0.09	-0.06	-0.33	-0.18	0.17	0.17	0.2	0.2
Row_4	Lab_4	0.33	0.1	0.07	0.29	0.21	0.46	0.46	0.43	0.43
Row_5	Lab_5	-0.67	-0.34	-0.23	-0.59	-0.69	0.24	0.24	0.37	0.37
Row_6	Lab_6	0.9	0.85	0.57	0.8	1.7	1.85	1.85	1.57	1.57
Row_7	Lab_7	-1.2	-0.27	-0.18	-1.06	-0.53	-0.03	-0.03	0.06	0.06
Row_8	Lab_8	-0.99	-0.34	-0.23	-0.87	-0.68	0.03	0.03	0.15	0.15
Row_9	Lab_9	0.99	0.65	0.43	0.88	1.3	1.88	1.88	1.57	1.57

Figure 14: Detailed statistics table (8 tests) from interlaboratory testing report

Suggestions

	Lab	Text
Row_2	Lab_2	Laboratory needs to proceed to corrective actions. These may include: a) checking that staff understand and follow t calibration of equipment and the composition of reagents d) replacing suspect equipment or reagents e) comparativ
Row_6	Lab_6	Laboratory Performance is unsatisfactory and a warning must be issued.
Row_9	Lab_9	Laboratory Performance is unsatisfactory and a warning must be issued.

Figure 15: Actions and warnings raised for particular labs in interlaboratory testing report.

Lab Real Names

	Given Name	Original Name %
Row_1	Lab_1	FCSV-8b479138-4775-3aba-b9cc-f01cc967d42b
Row_10	Lab_10	FCSV-8b479138-4775-3aba-b9cc-f33f3f3fnhhnyht6gbdsdwo9o
Row_2	Lab_2	FCSV-8b479138-4775-3aba-b9cc-fasdfasdfasdf
Row_3	Lab_3	FCSV-8b479138-4775-3aba-b9cc-f01certyertye
Row_4	Lab_4	FCSV-8b479138-4775-3aba-b9cc-f01cfghjfrty
Row_5	Lab_5	FCSV-8b479138-4775-3aba-b9cc-56j5j5j5j56
Row_6	Lab_6	FCSV-8b479138-4775-3aba-b9cc-f01w4thfh45jtjh
Row_7	Lab_7	FCSV-8b479138-4775-3aba-b9cc-f04hw24hh46u6i88l
Row_8	Lab_8	FCSV-8b479138-4775-3aba-b9cc-f344g4g4yj77j7j44
Row_9	Lab_9	FCSV-8b479138-4775-3aba-b9cc-f034334y3h3h3h3h77hjhkuiui

Figure 16: Concealed and real lab names dictionary in interlaboratory testing report

Finally, the report provides seven figures for visualisation purposes. Some of these are the histogram of estimated bias, normal probability plot of expanded uncertainties based on rank, normal probability plot of values with z-score annotations and plot of robust averages and standard deviations with significance values (Chi-square). These are visualised respectively in Figs (17-20).

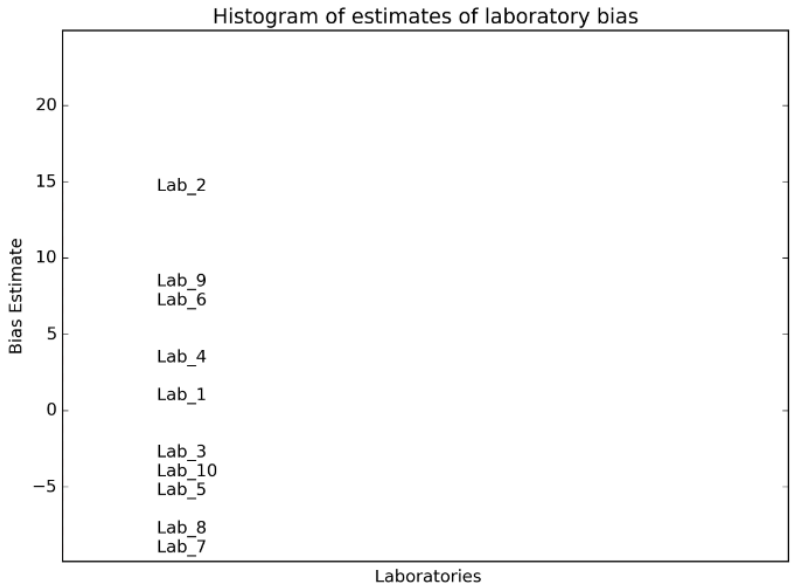


Figure 17: Histogram of bias estimation for bins of laboratories (auto-binning using Scott's method)

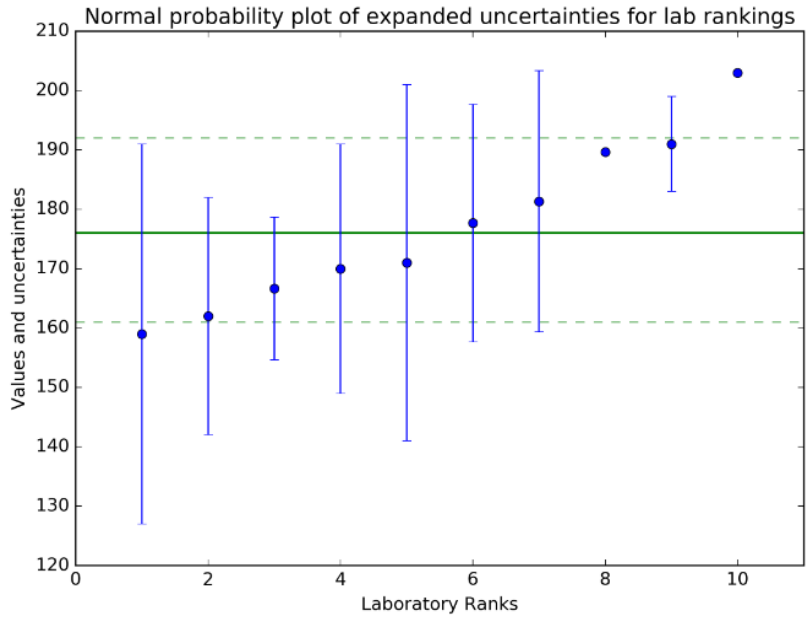


Figure 18: Normal probability plot of expanded uncertainties based on ranks

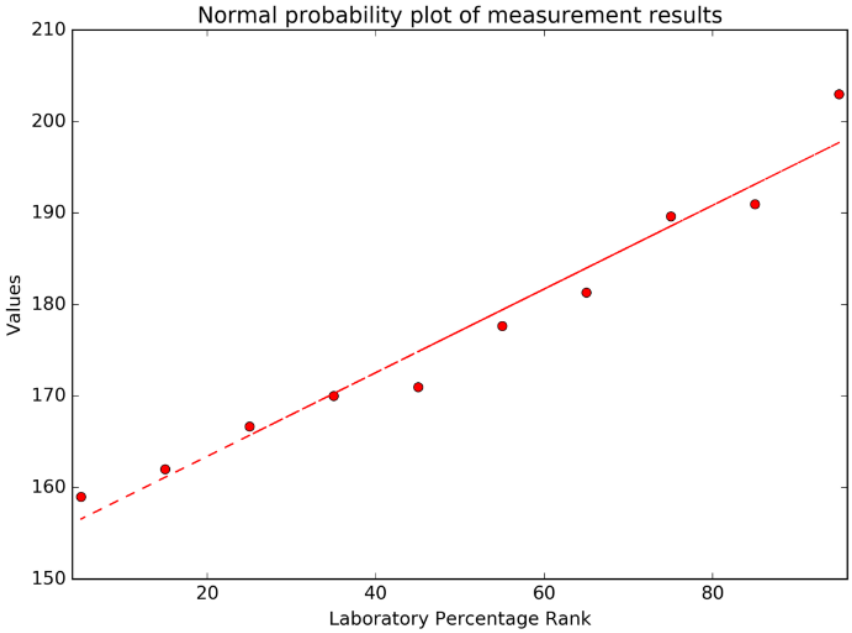


Figure 19: Normal probability plot of result values. Outliers outside Z-score threshold are annotated with their Z-score value.

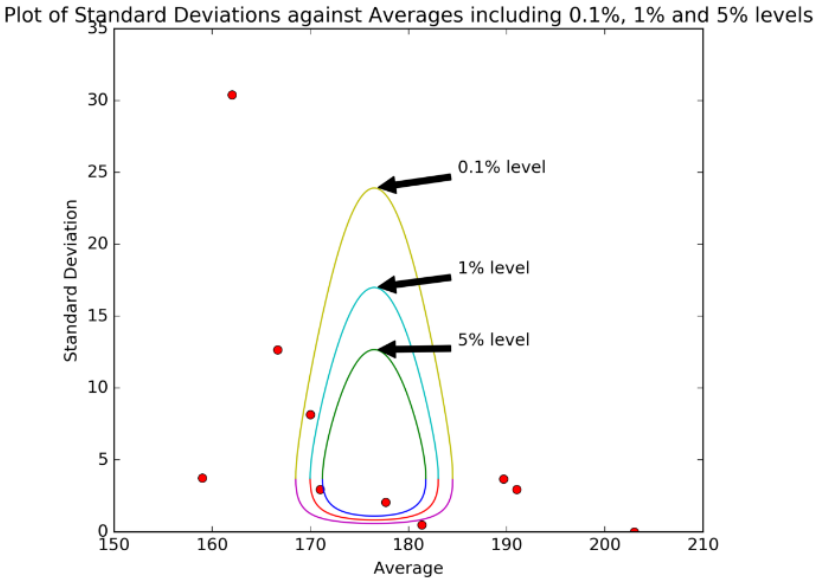


Figure 20: Plot of standard deviations against averages with significance values

5. CONCLUSION

In this report we have outlined the new functionalities added to JQ, namely optimal experimental design and interlaboratory proficiency testing. In the first case, experimental design functionalities for novel experiments or experiments in progress are introduced. The aim is to help the user to either design additional experiments of an ongoing research project by calling Iterative Experimental Design (IED) or initial experiments by calling the Factorial Experimental Design (FED) service. IED functionalities were implemented based on the iterative process described by Brandmaier et al. (2012) and Brandmaier and Tetko (2013), whereas FED functionality produces a classical full factorial design and suggests a set of trials given the selected optimality criterion. An important feature of both services is the interactivity of the dataset created, i.e. in all cases the user can add the new data collected (measurements of the endpoint variable) by following the instructions of the service and rerun the algorithm or proceed with other modelling functionalities. In the second case, interlaboratory testing (ILT) will allow researchers to assess laboratory bias in a straightforward fashion (a matter of several 'clicks') and proceed to corrective actions in case of suboptimal performance. This was implemented according to ISO standards for consensus opinion using rigorous statistical testing, and allows the ILT coordinators to preserve the anonymity of all laboratories when providing them with the outcomes, should this be desired.

6. BIBLIOGRAPHY

Atkinson, AC; Donev, AN; Tobias, RD. Optimum Experimental Designs with SAS. *Oxford University Press* 2007.

Brandmaier, S; Tetko, IV. Robustness in experimental design: A study on the reliability of selection approaches. *Comput Struct Biotechnol J,* 2013, 7(9): 1-10, DOI: 10.5936/csbj.201305002.

Brandmaier, S; Sahlin, U; Tetko, IV; Öberg, T. PLS-Optimal: A stepwise D-Optimal design based on latent variables. *J Chem Inf Model*, 2012, *52*(4), 975-983, DOI: 10.1021/ci3000198.

Dette, H; Bretz, F; Pepelyshev, A; Pinheiro, JC. Optimal Designs for Dose Finding Studies. *J Amer Statist Assoc*, 2008, 103 (483), 1225–1237, *DOI*: 10.1198/016214508000000427.

Federov, VV. Theory of optimal experiments, Academic Press NY 1972.

Wheeler, RE. Comments on algorithmic design. *Vignette accompanying package AlgDesign* 2004.

ISO 13528, Statistical methods for use in proficiency testing by interlaboratory comparison, 2015.