NanoCommons

Nano-Knowledge Community

The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

Grant Agreement No 731032

Deliverable Report 5.3

Deliverable	D5.3. First raw data processing and knowledge extraction tools integrated into KB and first theoretical descriptor calculation services available
Work Package	WP5: JRA3 - Analysis and Modelling tools
Delivery date	M12 - 31 December 2018
Last update	(writing)
Lead Beneficiary	National Technical University of Athens (NTUA)
Nature of Deliverable	Demonstrator
Dissemination Level	Public

Submitted by	Philip Doganis, Haralambos Sarimveis, Angelos Valsamis (NTUA), Antreas Afantitis, Georgia Melagraki, Dimitra-Danai Varsou, Andreas Tsoumanis (NovaM), Konstantinos Kotsis and Vladimir Lobaskin (UCD)
Revised by	Stacey Harper (UO) and Antreas Afantitis (NovaM)
Approved by	Iseult Lynch (UoB)

Table of contents

Abbreviations

API: Application Programming Interface
CDK: Chemistry Development Kit
CG: Coarse Grained
CIS: Configuration Interaction Singles
CNT: Carbon NanoTubes
COSMO: Conductor-like Screening Model
DFT: Density Functional Theory
DFTB: Density Functional Tight-Binding
DNA: Deoxyribo Nucleic Acid
ESPResSo: Extensible Simulation Package for Research on Soft Matter
FAIR: Findable, Accessible, Interoperable and Re-usable
GUI: Graphical User Interface
GPU: Graphics Processing Unit
GROMACS: GROningen MAchine for Chemical Simulations
HCNP: Hydrophobic Charged Nanoparticles
HF: Hartree-Fock
iTasser: Iterative Threading Assembly Refinement
JSON: JavaScript Object Notation
KB: Knowledge Base
LDM: Liquid Drop Model
MD: Molecular Dynamics
MO: Metal oxides
MPI: Message Passing Interface
nanoQSA/PR: nano Quantitative Structure Activity/Property Relationship
NM: Nanomaterial
NP: Nanoparticle
NWChem: North West Chemistry
PDB: Protein Data Bank
PM6 or PM7: Parametrization Method 6 or 7
QSAR: Quantitative Structure Activity Relationships
REACH: Registration, Evaluation and Authorisation of Chemicals
SDF: Spatial Data File
SEM: Scanning Electron Microscopy
SMILES: Simplified Molecular-Input Line-Entry System
SNT-MT: SmartNanoTox Modelling Tools
SOAP: Simple Object Access Protocol
TA: Transnational Access
TEM: Transmission Electron Microscopy
TDDFT: Time-Dependent Density Functional Theory
TM: Transition metals
UI: User Interface
URI: Uniform Resource Identifier
vdW: van der Waals
WSDL: Web Services Description Language

Summary

This Deliverable (D5.3) focuses on (i) tools developed for processing of raw data related to nanomaterials and their safety, and their valorisation for extracting additional knowledge to enrich datasets and facilitate modelling and risk assessment; and (ii) tools for the calculation of theoretical descriptors such as molecular or atomic computational descriptors, that are at a stage of development suitable for implementation into the NanoCommons toolbox as services available for Transnational Access (TA).

The tools for processing raw data focus on nanoparticle (NP) or nanomaterial (NM) image analysis, with the raw data being sourced from Scanning or Transmission Electron Microscopy (SEM/TEM) images, in order to extract additional quantitative and qualitative descriptors beyond the usual size, size distribution, and a qualitative description of shape and agglomeration. The tools provide a more accurate, complete and streamlined characterisation procedure and increase the amount of quantitative information on nanomaterials (NM) that can be used for modelling and improved understanding thereon. Integration of the data extracted using these tools directly into the NanoCommons KnowledgeBase ensures ease of implementation and streamlining of services available to TA Users, as well as supporting NanoCommons's goal to enhance the FAIRness and Openness of datasets in order to facilitate further re-use of the data within the nanosafety and material modelling communities.

Theoretical descriptors are produced through two different routes: First, cheminformatics open-source software CDK (Chemistry Development Kit) produces descriptors from SMILES electronic molecular structure representations. Second, atomistic and coarse-grained Molecular Dynamics (MD) calculations, performed with a variety of MD packages, are used to evaluate the material descriptors including intrinsic and extrinsic NM properties (electronic structure characteristics, reactivity, hydration energy, van der Waals energies etc.), as well as quantitative descriptors of bionano interactions (biomolecular adsorption energies and binding affinities) that provide as basis for prediction of protein corona formation and composition.

Introduction

Efficient characterization of nanomaterials (NM) is critical in reviewing and assessing their value, suitability and safety, either as end-products or for specific use cases. Proper collection, formatting and integration of characterisation information is essential for developing modelling infrastructure for computationally predicting possible adverse effects, and consequently building a framework for risk assessment, risk management and for designing safer NMs. Characterisation of NMs is not a trivial task, as it needs to process, curate and combine the results of a wide range of experimental tests for the derivation of raw physicochemical and biological experimental data, some of which can then be utilised for the calculations of theoretical descriptors. Additionally, purely computational or theoretical descriptors, based on the structure of NMs, can also be determined further enriching the available data describing NMs. Indeed, the EU -US Nanoinformatics roadmap identifies two types of descriptors as follows: “Descriptors can be experimentally measured properties, usually related to the physical or chemical identity of NMs, and theoretical descriptors, which are derived from the electronic, atomistic and molecular structure of NMs and their immediate environment.”

In this deliverable we describe and demonstrate the tools and services that have been developed during the first year of the NanoCommons project for processing and extracting knowledge (image-derived descriptors of NMs physicochemical properties and/or degree of agglomeration / aggregation) from raw experimental data and for calculating purely theoretical descriptors that may or may not relate to experimentally measurable features. The deliverable summarises the work and developments performed so far in T5.1 and T5.3 of the project. The first section of this deliverable demonstrates the tools developed for processing and knowledge extraction from electronic NM images, which are extremely valuable sources of information that are currently not exploited at all in nanosafety assessment beyond a size distribution and a qualitative description of shape.   The second section focuses on the tools that can be used for the calculation of purely theoretical descriptors, that don’t require any experimental input but rather utilise information on the unit cell or crystal structure of the core NM. Combined with the tools developed in T5.2 and described in D5.2 which focuses on the analysis and processing of biological data, NanoCommons is building a complete, integrated and harmonised framework for the characterisation of NMs, linked directly to the NanoCommons Knowledge Base (enriched database) and ontologies.

Description

Analysis of raw experimental data contained in electronic images

The features and morphology of NMs affect their properties to a great extent and can be used as input information for the development of nano Quantitative Structure Activity/Property Relationship (nanoQSA/PR) models that predict properties and/or adverse biological/environmental effects. NM characterization immediately following synthesis is routinely performed to confirm that the desired product was produced. One gold-standard characterization method is transmission electron microscopy (TEM) whereby images are obtained to determine quantitatively the size (by measuring large numbers of particles) and qualitatively the shape of the NMs and other morphological characteristics (Mondini et al., 2012).

Electronic images are invaluable sources of information. Electronic image data files contain big raw data that need to be further processed in order to extract useful knowledge for the characterisation of NMs. From an *in silico* point of view, a range of different variables, the so-called image descriptors, can be extracted from a given NM TEM image with the use of appropriate software. These image descriptors can then be used to develop meaningful correlations with an activity or property of interest. Past examples of such efforts, where image descriptors have been calculated and then used to derive nanoQSA/PR models, have been reported in the literature (Mikolajczyk et al., 2015 and Toth et al., 2013). However, these tools have not been made publically available or linked to knowledge to management infrastructure, limiting their usefulness. Another aspect of the greater success that open source algorithms demonstrate is recognising NMs in biological environments, such as within a cell, in comparison to the commercial packages that accompany electronic microscopes. The algorithms presented here are available through ImageJ, among other tools which aligns with current practice for many researchers that involves taking microscope images and then analyzing them in ImageJ. A web-based solution that provides (i) a User Interface to easily test the capabilities of an image analysis tool, (ii) complete access to its calculations, and (ii) easy integration to existing infrastructures through the use of web services, is an important contribution to academic research, as well as facilitating regulatory validation of these open-source tools as being suitable for use in NMs registration dossiers under EU chemicals legislation (EC 1907/2006) REACH.

Manual image characterization is most commonly used to extract characterisation features, but it is time consuming, it is exposed to subjectivity and might have variability in the measurements when performed by different people or even by the same specialist. Automatic image analysis has the potential to overcome the disadvantages of manual methods. It reduces the time and effort required for NM characterization and does not require observer intervention, overcoming subjectivity problems and finally, it can help the creation of versatile tools to work on images acquired by any microscope, unlike current specialized programs. Additionally, automated image analysis coupled with machine learning may, in the future, be able to detect subtle changes in agglomeration state or ageing-related changes that are not discernible to the human eye.

The NanoCommons infrastructure offers two image analysis tools, namely the NanoImage tool developed by partner NTUA and the NanoXtract tool developed by partner NovaM. These complementary tools are presented with Graphical User Interfaces (GUIs), and are additionally offered as web-services to allow their integration into modelling schemes and their widespread utilization by the nanosafety and NMs synthesis and materials modelling communities. In the rest of this section the tools are described and demonstrated with example electronic images.

NanoImage - Jaqpot Image Analysis Web Application (NTUA)

During the eNanoMapper FP7 project, the first version of the Jaqpot image analysis application was developed by NTUA, focusing on spherical particles. It was one of the first implementations that brought the functionality of ImageJ (Rueden et al., 2017) from desktop computers to the web as a web application, allowing users to perform image analysis on a browser without any software installation, accessible from PCs or even mobile devices. New developments in Jaqpot architecture and available tools made it necessary to update its code in order to make its inclusion in the new NanoCommons TA application possible. In the context of the NanoCommons project, the NanoImage tool has been extended by including functionalities to process electronic Carbon NanoTubes (CNT) images. The tool is released with a Graphical User Interface (GUI) and is also available as a web service, which allows its integration into the NanoCommons infrastructure. The user interface is hosted at the following address: https://app.jaqpot.org/nanoImage/. The web services are accessible through the Jaqpot Swagger documentation: https://api.jaqpot.org/jaqpot/swagger/. The source code of the application is available at: https://github.com/KinkyDesign/image-analysis, so that the tools can be further progressed through work of the community and can be adapted to specific user needs (for example analysing similarity of different batches of NMs, estimating ageing etc.) via user applications for Transnational Access (TA), for example.

NanoImage - Image Descriptor calculations for spherical NMs

Given an electronic image containing NMs of spherical shapes the NanoImage tool offers the option of applying different thresholding filters (*Huang* (Huang & Wang, 1995), *Triangle* (Zack et al. 1997), Intermodes (Prewitt & Mendelsohn, 1966), *Isodata*, *IJ_IsoData* (Ridler & Calvard, 1978), *Li* (Sezgin & Sankur, 2004), *MaxEntropy* (Kapur et al., 1985), *Mean* (Glasbey, 1993), *Moments* (Tsai, 1985), *Otsu* (Otsu, 1979), *RenyiEntropy* (Kapur et al., 1985), *Shanbhag* (Shanbhag, 1994)). The tool is then able to identify the different NMs contained in the electronic image and automatically compute the descriptors presented in Table 1:

Table 1: NanoImage - Calculated Image descriptors for Spherical nanoparticles.

Image descriptor	Brief meaning	Range of values
Angle	The angle between the primary axis of the best fitting ellipse and a line parallel to the x-axis of the image.	0-180 [deg]
Aspect Ratio	The quotient of the division: *major_axis/minor_axis*.	>0 unitless
Circularity	The degree to which a NP approaches a perfect circle.	0-1 [unitless]
Feret Angle	The angle between the (min) Feret diameter and a line parallel to the x-axis of the image.	0-180 [deg]
Grey deviation	The deviation of the grey values used to count the mean grey value.	>0

Integrated Density	The product of Area and Mean Gray Value.	>0
Kurtosis	The fourth order moment about the mean of gray values.	[unitless]
Major axis/ Major	The longest diameter of the best fitting ellipse to the NP.	>0 [nm]
Max Grey Value	The highest grey scale values appearing in the selection.	>0
Mean Grey Value	The average grey value within the selection.	>0
Min Grey Value	The lowest grey scale values appearing in the selection.	>0
Minimum Feret's diameter/ Min Feret	The shortest distance between any two points along the selection boundary.	>0 [nm]
Minor axis/ Minor	The shortest diameter of the best fitting ellipse to the NP.	>0 [nm]
Modal Grey Value	Stands for the most frequently occurring grey value within the selection.	>0
Perimeter	Total length of the NP's boundary.	>0 [nm]
Porosity	Provides a measure of porosity.	>0
Roundness	Compares the surface of the NP to the surface of the disc of diameter equal to major axis.	0-1 [unitless]
Size/Area	The area of the NP.	>0 [nm2]
Skewness	The third order moment about the mean of gray values.	[unitless]
Solidity	The degree of the overall concavity or convexity of a NP.	0-1 [unitless]
Sphericity	The ratio of the surface area of a sphere to the surface area of the particle.	>0 unitless
Surface diameter	The diameter of a sphere having the same surface area as the projected particle.	>0 [nm]
Volume	Estimation of the volume of the NM	>0 [nm3]
Volume to Surface	The diameter of a sphere having the same volume to surface ratio as the projected particle.	>0 [nm]

An explanation of the above-mentioned terms, their units and their correspondence to the descriptors calculated by NanoXtract are provided in Appendix 2, while their ontological annotation is provided in Appendix 4.

The results are presented in the form of a table which contains the values of the selected descriptors for each particular NM and the averages over all NMs. A detailed description was provided in Deliverable Report D4.2 of eNanoMapper (Doganis et al. 2015), which will be integrated into the

NanoCommons TA website. Users are encouraged to consult it for in-depth descriptions.

The application has been extended within NanoCommons in relation to its development stage eNanoMapper in the following aspects:

- The following thresholding filters have been added: (Li 4, MaxEntropy, Mean, Moments, Otsu, RenyiEntropy, Shanbhag). Since the result of a thresholding algorithm is difficult to evaluate automatically, we made available almost all of the thresholding algorithms that the ImageJ Auto Threshold plugin (Schindelin et al., 2017) offers. The end-user can easily test filters on the initial image and evaluate the quality of the extraction empirically.
- Export feature, which allows export of the resulting table in various formats (.pdf, .xls, .csv, .xml) for further use and integration of the results into other data processing and modelling applications, including those offered by NanoCommons.
- Pixel to nm ratio: this provides the ratio between each pixel in the image and its actual dimensions in nanometers, leading to results measured in nm which is the standard experimental unit or NM particle size.

NanoImage - Image Descriptor calculations for CNTs

In the context of the NanoCommons project the NanoImage tool was extended to cover the big family of Carbon Nanotubes (CNTs). The characterization process of CNTs aims to determine several features, that may include chemical composition, surface chemistry and morphology, randomness or alignment of CNTs growth, as well as their orientation. The CNT component of the NanoImage analysis application is based on making the functionality of the Ridge detection plugin (Steger, 1998) available over a browser. The CNT descriptors that are determined are: *Angle of Normal*, *Countour Class*, *Contrast*, *Asymmetry*, *Line Width*, *Estimated Length* and are briefly presented in Table 2. They are described in more detail in Appendix 3, while the ontology terms they correspond to are reported in Appendix 4.

Table 2: NanoImage - Calculated Image descriptors for CNTs

Image descriptor	Brief meaning	Range of values
Angle of Normal	The direction of the normal to each line point.	[rad]
Asymmetry	The true contrast of each line point.	
Contour Class	Characterise the CNT line (lines are called contours in the plugin) regarding presence and type of junctions.	
Contrast	The asymmetry of each line point.	

(Mean) Line Width	**Line Width** refers to CNT section and equals to the sum (WidthL +WidthR). **Mean Line Width** refers to entire CNT and equals to the mean value of this feature along the CNT.	>0 [nm]
(Estimated) Length	These descriptors provide the length either of the CNT section, or the entire CNT, respectively.	>0 [nm]

This plugin was selected among the plethora of available plugins after meticulous testing of the available plugins on electron microscopy images was performed and the efficacy of their results was compared to the experimental measurements taken by a laboratory scientist. The Ridge detection plugin exhibited the best behavior in the recognition of nanotubes with varying curvature and widths and provided measurements that were close to the experimental results, whilst also providing an intuitive overlay on the original image, easily showing the nanotubes recognised. Additional advantages of this plugin were:

- the few parameters that are required to derive results;
- the relatively easy understanding of the input parameters' role in modulating the outputs from the image analysis;
- the lack of need for filter selection by the user.

Demonstration of the NanoImage Graphical User Interface

The workflow of the web application starts with the landing page https://app.jaqpot.org/nanoImage/ (Figure 1), which contains information about the tool (Figure 2) and an expandable menu allowing the user to navigate among the two applications, namely spherical particles and nanotubes (Figure 3).

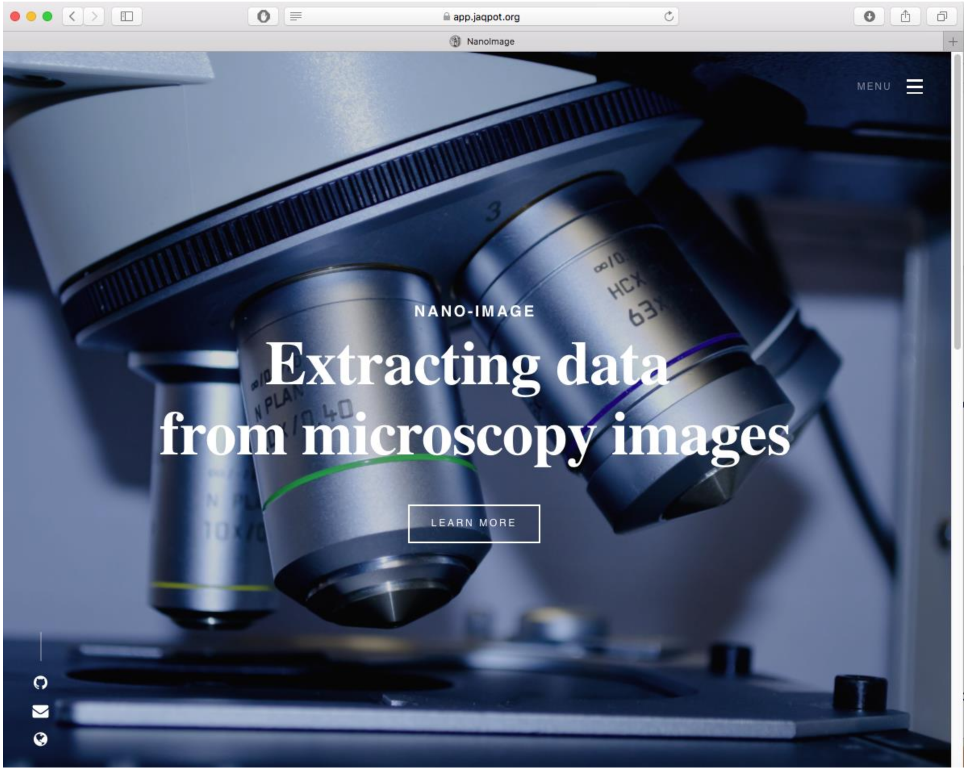


Figure 1. The landing page of the NanoImage GUI

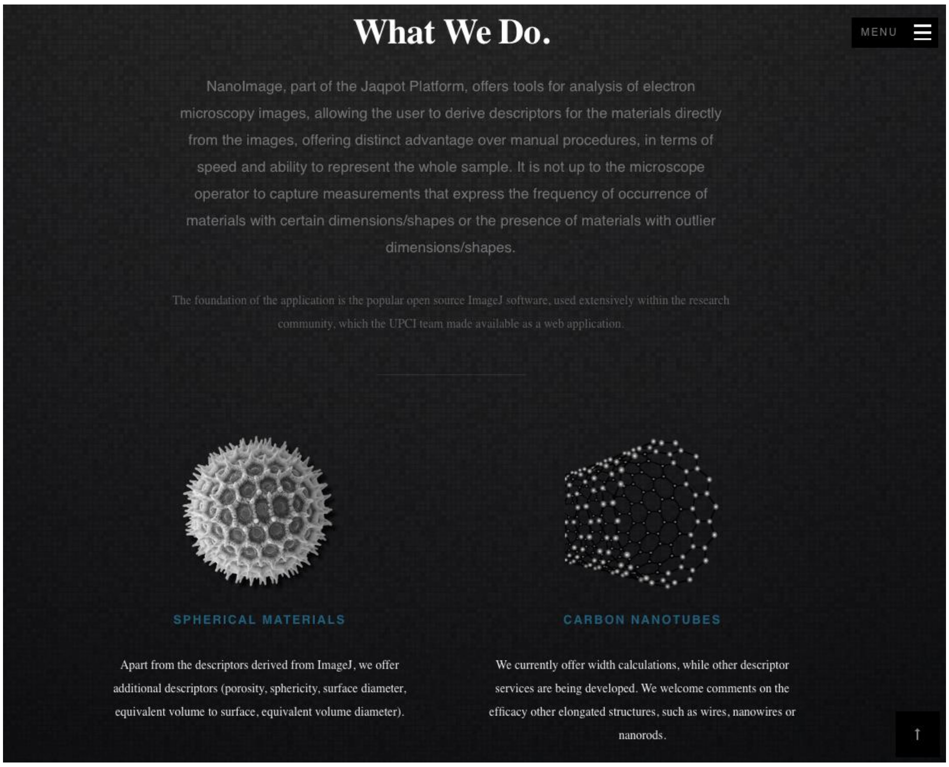


Figure 2. Brief presentation of NanoImage functionality on the landing page

Figure 3. NanoImage GUI landing page - Navigation menu

Spherical Particles

In the case of spherical particles, the analysis of electron microscopy images with NanoImage begins by uploading the image (Figure 4). Users should click on the button with the text “Choose a file from the system” and select the image in the folder of their computer and click Insert. Alternatively, users can test the application using an example image. The test image is derived from the publication by Walkey et. al. (2014), which studied the protein corona of gold nanoparticles. More specifically, results on 105 gold nanoparticles showed the protein corona to produce more accurate predictions of cell association than parameters involving nanoparticle size, surface charge and possible aggregation state. The test image depicts gold nanoparticles (without a corona) and was derived using Transmission electron microscopy (TEM).

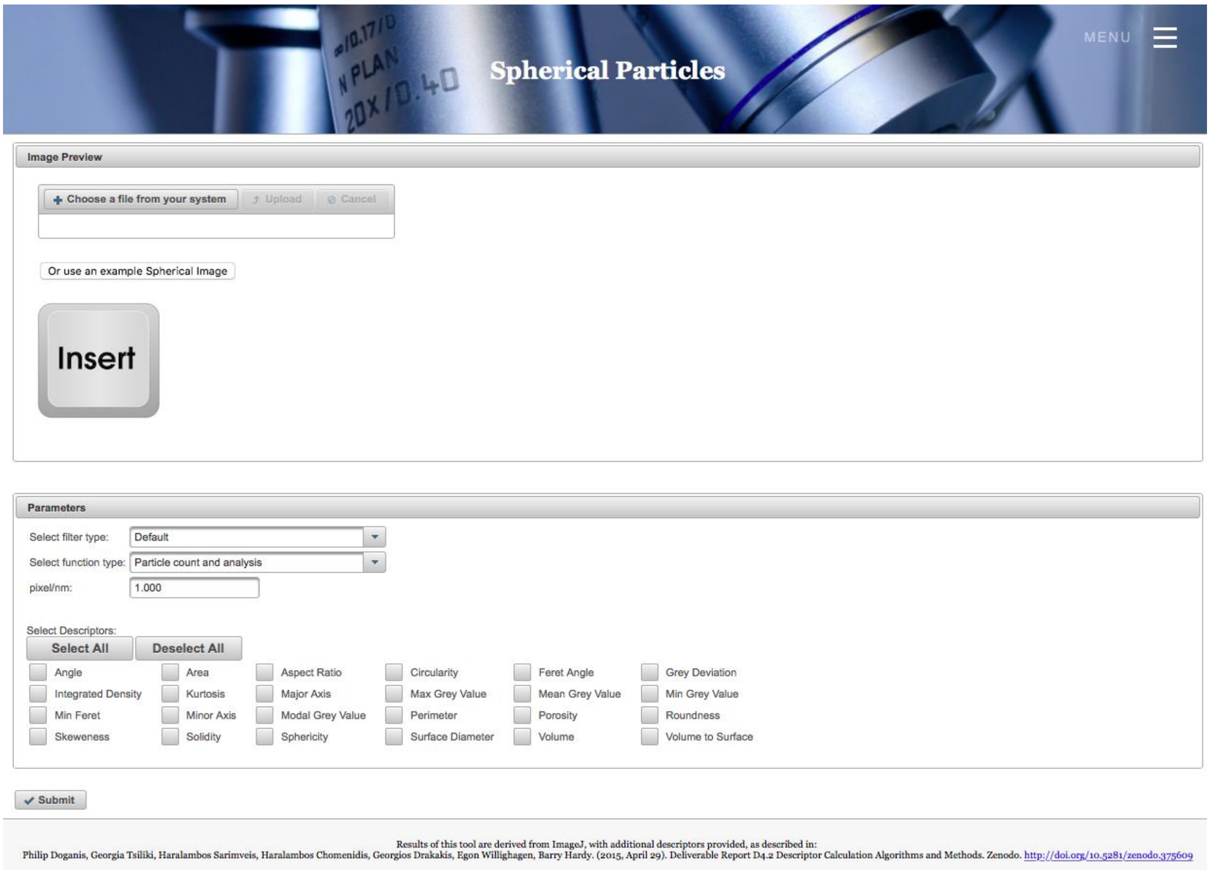


Figure 4. NanoImage: Spherical particles descriptor calculations

This leads to users seeing the preview of the image after the default filter has been applied and a red overlay defines the areas that are separated from the background and recognised as nanoparticles (Figure 5). The user then needs to select the following parameters:

- Select filter type (Figure 6): This can be: Default filter (named *default* in ImageJ, a variation of the IsoData algorithm), Huang, Triangle, Intermodes, Isodata, IJ_IsoData, Li, MaxEntropy, Mean, Moments, Otsu, RenyiEntropy, Shanbhag.

  Its default value is the Default filter.

- Select function type (Figure 7):
  - Full image Analysis

    All areas in the image depicting nanomaterials are regarded as one by the algorithm, producing cumulative statistics that correspond to one hypothetical large particle.

  - Particle count and analysis

    Each particle is measured individually and results are provided for both each particle and average values that represent the entire population.

- Select Descriptors (Figure 7):

    Users select the descriptors they desire to be calculated. Besides the Select all/Deselect all options, it is also possible to make a go/not-go selection for each one of these descriptors: *Angle, Area, Aspect Ratio, Circularity, Feret Angle, Grey deviation, Integrated Density, Kurtosis, Major Axis, Max Grey Value, Mean Grey Value, Min Feret, Min Grey Value, Minor Axis, Modal Grey Value, Perimeter, Porosity, Roundness, Skeweness, Solidity, Sphericity, Surface Diameter, Volume, Volume to Surface*. Appendix 4 gives a brief description of each of these descriptors aligned with the NanoCommons ontology.

- Pixel to nm ratio (Figure 5):

    The ratio of pixels per nm is provided here by the user and is a meta information contained in the image file.

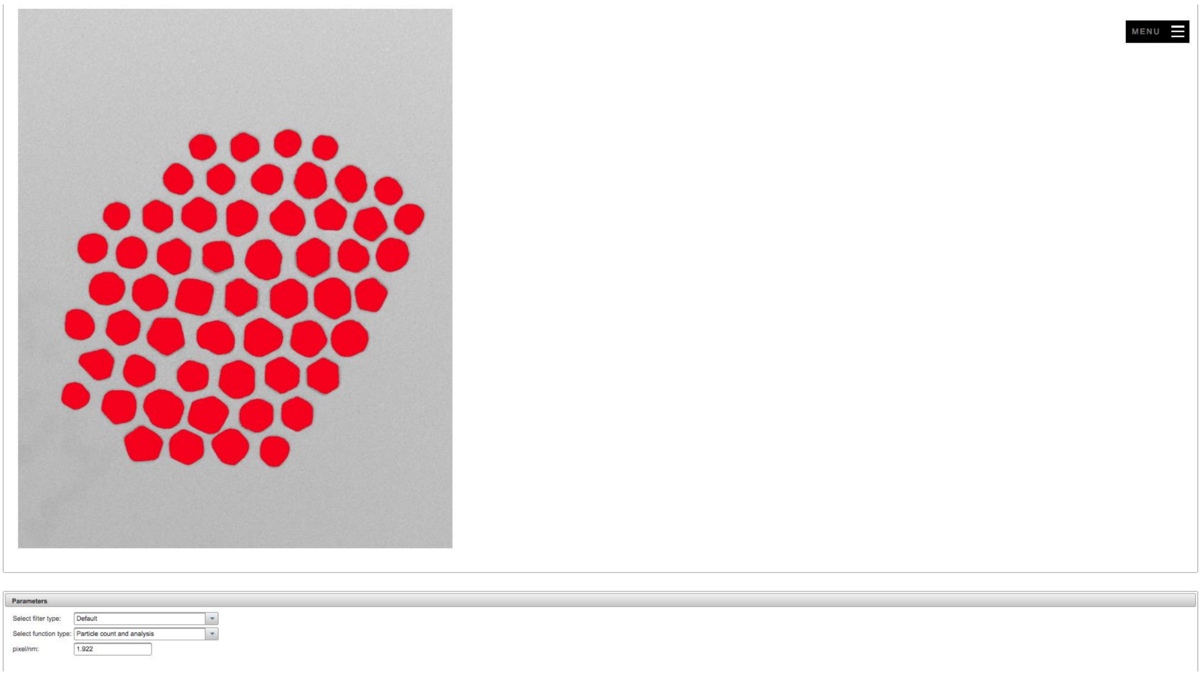


Figure 5. NanoImage: user input page for Spherical particles

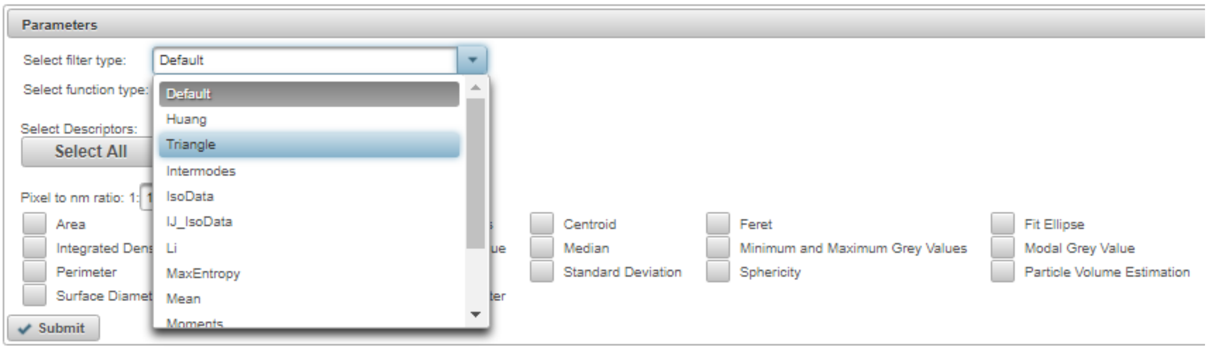


Figure 6. NanoImage: filter selection for Spherical particles

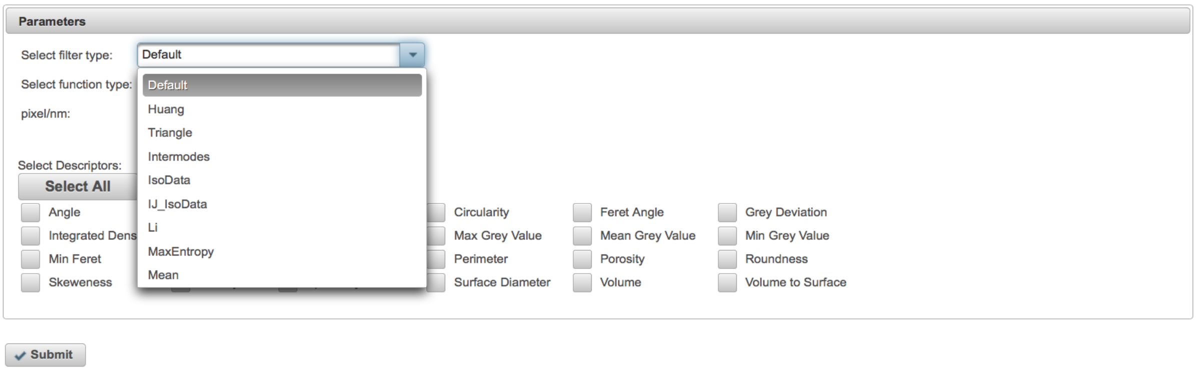


Figure 7. NanoImage: function type selection for Spherical particles (analysis of individual particles or entire surfaces).

After clicking submit, the user is directed to the results page that provides a preview of the image analysed, where the recognised areas are marked in order to make the basis of the calculations clear. In Figure 8, the results page for the Particle count and analysis function is provided as an example. The average values for all descriptors calculated are provided in the first row, named ”Average particle”. Measurements for the individual particles are provided in the following rows. Besides viewing the results on the webpage, it is possible to export and download them in various formats (.pdf, .xls, .csv, .xml), as shown in Figure 9.

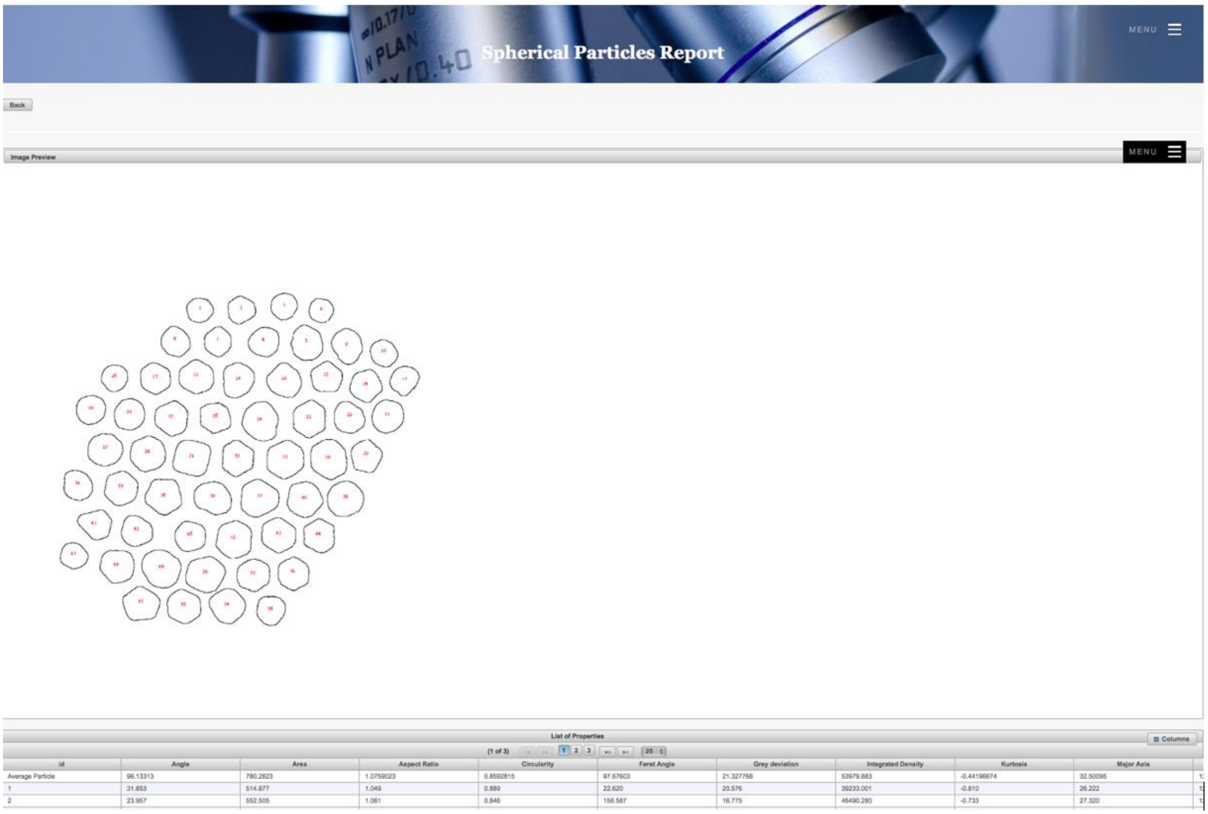
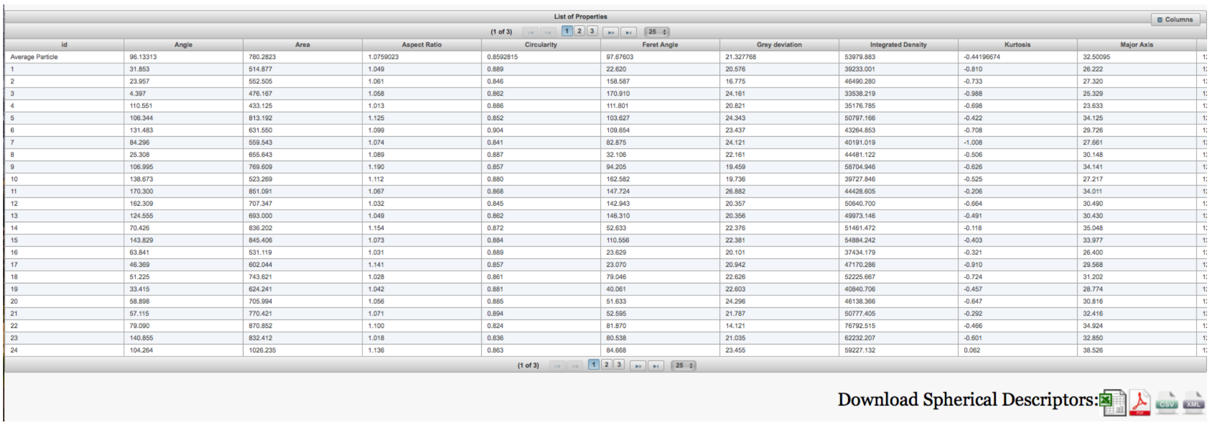


Figure 8. NanoImage: Result page for Spherical particles



List of Properties

id	Angle	Area	Aspect Ratio	Circularity	Feret Angle	Grey deviation	Integrated Density	Kurtosis	Major Axis
Average Particle	96.13313	780.2823	1.0759023	0.8592815	97.67603	21.327768	53979.883	-0.44196674	32.50095
1	31.853	514.877	1.049	0.889	22.620	20.576	39233.001	-0.810	26.222
2	23.957	552.505	1.061	0.846	158.587	16.775	46490.280	-0.733	27.320
3	4.397	476.167	1.058	0.862	170.910	24.161	33538.219	-0.988	25.329
4	110.551	433.125	1.013	0.886	111.801	20.821	35176.785	-0.698	23.633
5	106.344	813.192	1.125	0.852	103.627	24.343	50797.166	-0.422	34.125
6	131.483	631.550	1.099	0.904	109.654	23.437	43264.853	-0.708	29.726
7	84.296	559.543	1.074	0.841	82.875	24.121	40191.019	-1.008	27.661
8	25.308	655.643	1.089	0.887	32.106	22.161	44481.122	-0.506	30.148
9	106.995	769.609	1.190	0.857	94.205	19.459	58704.946	-0.626	34.141
10	138.673	523.269	1.112	0.880	162.582	19.736	39727.846	-0.525	27.217
11	170.300	851.091	1.067	0.868	147.724	26.882	44428.605	-0.206	34.011
12	162.309	707.347	1.032	0.845	142.943	20.357	50640.700	-0.664	30.490
13	124.555	693.000	1.049	0.862	146.310	20.356	49973.146	-0.491	30.430
14	70.426	836.202	1.154	0.872	52.633	22.376	51461.472	-0.118	35.048
15	143.829	845.406	1.073	0.884	110.556	22.381	54884.242	-0.403	33.977
16	63.841	531.119	1.031	0.889	23.629	20.101	37434.179	-0.321	26.400
17	46.369	602.044	1.141	0.857	23.070	20.942	47170.286	-0.910	29.568
18	51.225	743.621	1.028	0.861	79.046	22.626	52225.667	-0.724	31.202
19	33.415	624.241	1.042	0.881	40.061	22.603	40840.706	-0.457	28.774
20	58.898	705.994	1.056	0.885	51.633	24.296	46138.366	-0.647	30.816
21	57.115	770.421	1.071	0.894	52.595	21.787	50777.405	-0.292	32.416
22	79.090	870.852	1.100	0.824	81.870	14.121	76792.515	-0.466	34.924
23	140.855	832.412	1.018	0.836	80.538	21.035	62232.207	-0.601	32.850
24	104.264	1026.235	1.136	0.863	84.668	23.455	59227.132	0.062	38.526

Download Spherical Descriptors:

Figure 9. NanoImage: Output formats available - Results page for Spherical particles

Carbon Nanotubes

For analysis of CNTs, users are directed to a different input page, customised for features of nanotubes and the options to be made in that case (Figure 10). As before, users provide their image or use an example image and click insert. This leads to a preview of the image appearing on the page (Figure 11). The example image is shown in Figure 12. The image is availabe at https://upload.wikimedia.org/wikipedia/commons/9/9b/CNTSEM.JPG, by Materialscientist, License: Creative Commons Attribution-Share Alike 3.0 Unported

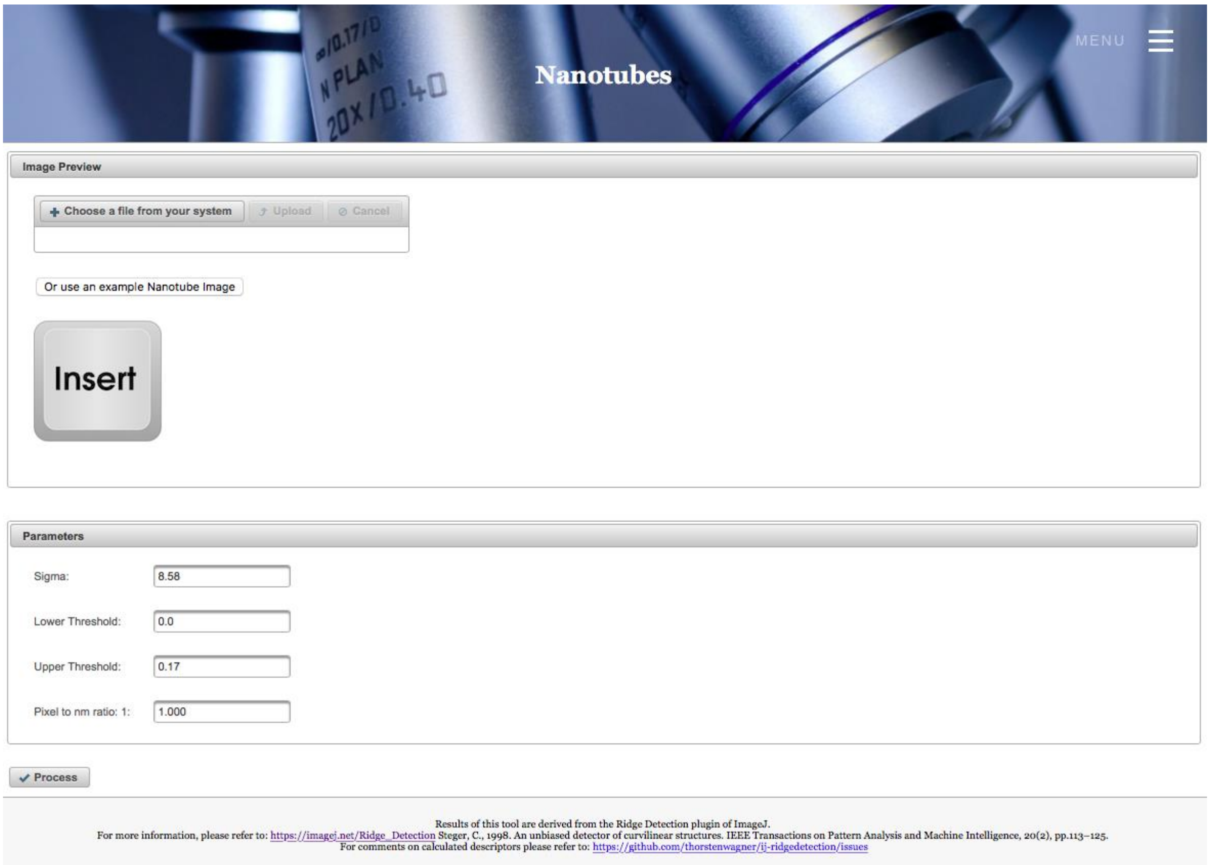


Figure 10. NanoImage: Nanotubes descriptor calculations

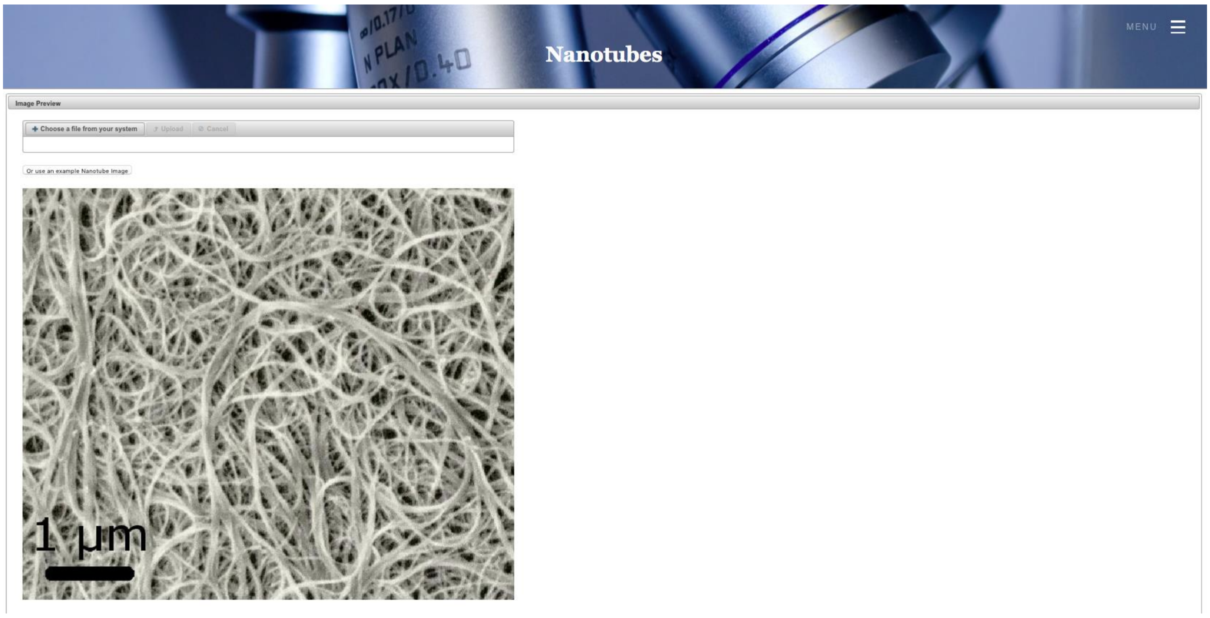


Figure 11. NanoImage: user input page for nanotubes

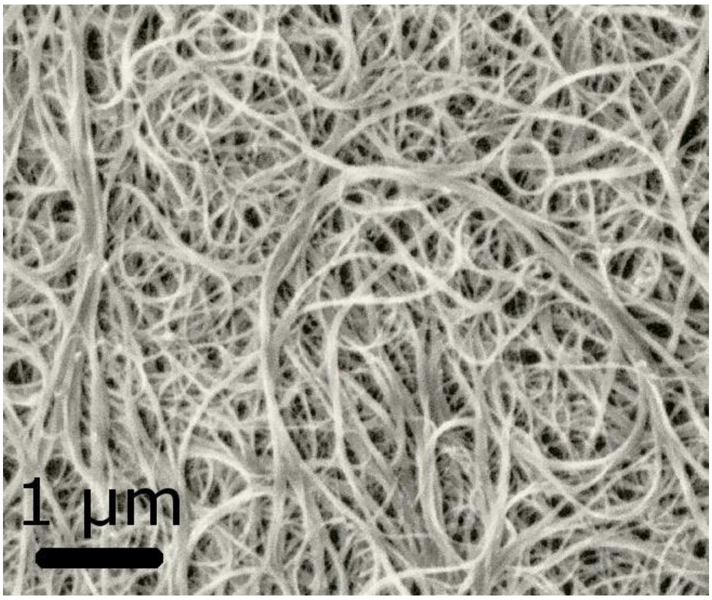


Figure 12. SEM image of carbon nanotube bundles

Additional mandatory parameters are required, as defined by the ImageJ plugin Ridge Detection (Steger, 1998), available at https://imagej.net/Ridge_Detection:

- **Sigma:** Determines the sigma for the derivatives. It depends on the line width in the sense that it can only be used to detect lines with a certain range of widths, i.e., between 0 and 2.5 σ. It must be noted that if σ is selected so large that neighboring lines start to influence each other, the line model will not be recognised effectively and the results will deteriorate. The following relation between width and sigma has to hold (Eq. (10), (Steger, 1998)):

$$\sigma \geq \frac{w}{\sqrt{3}}$$

- **Lower Threshold:** Line points with a response *smaller* than this threshold are rejected
- **Upper Threshold**: Line points with a response *larger* than this threshold are rejected.

Clicking on the Process button leads to the results page, shown in Figure 13, where users are provided with a preview of the processed image, allowing them to view the nanotubes that were recognised as an overlay on the original image for easier understanding and monitoring of performance. A histogram of the distribution of Mean Line Width (Nanotube thickness) measurements is provided below the images, as shown in Figure 14. Measurements are grouped into dynamically sized bins.

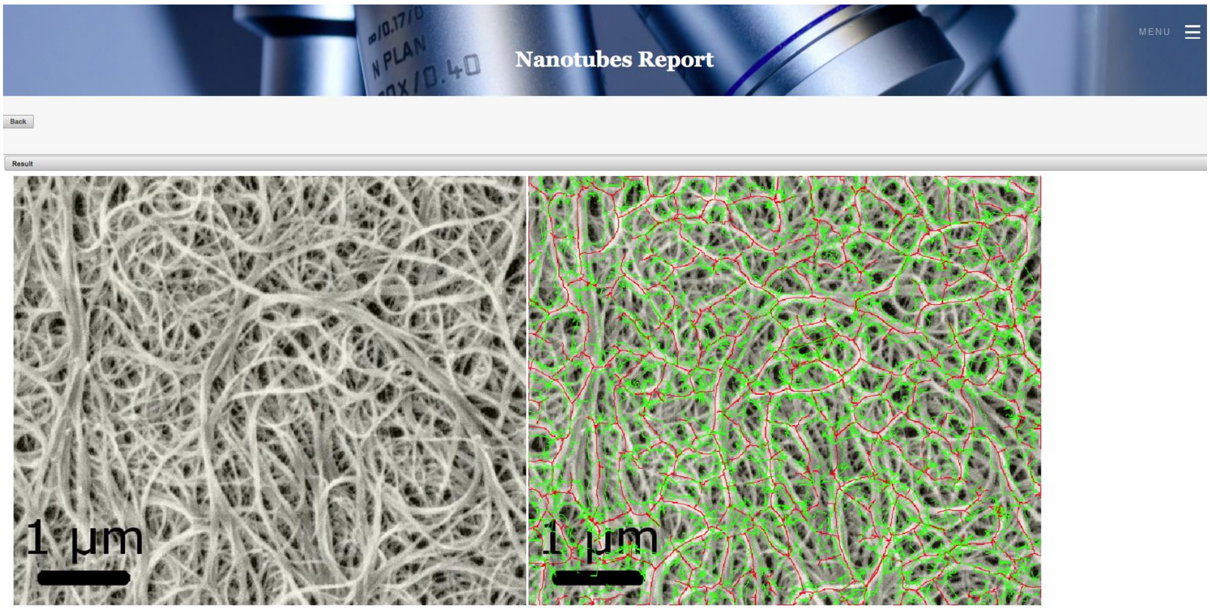


Figure 13. NanoImage: Result page for Nanotubes

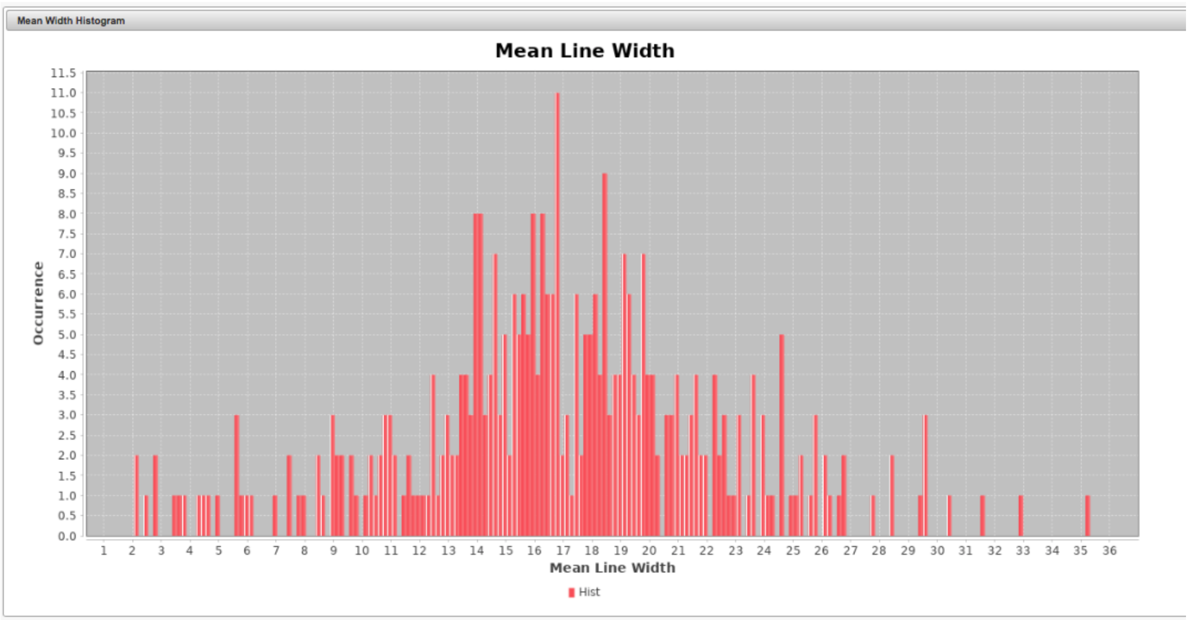


Figure 14. NanoImage: Distribution of Mean Line Widths among nanotubes identified in the image

The measurements for each nanotube, identified by contour id (a unique number for archiving purposes assigned to each CNT recognised as a contour), namely Length and Mean Line Width are provided in a table below the histogram (Figure 15). The first row contains the average values over all measurements made, with subsequent rows being the values for individual tubes. Each nanotube is represented by one row in this table. The results can be exported by downloading them in various formats. Below the table are the links that provide this functionality (Figure 15).

Summary per Nanotube

(1 of 18) 1 2 3 4 5 6 7 8 9 10 25

Contour Id.	Length	Mean Line Width
Average	48.454	19.366
1898	19.418	28.483
1879	17.790	24.792
2763	7.478	25.183
2149	46.242	25.178
2294	14.275	16.801
3027	32.286	12.935
1903	14.343	26.992
2919	13.759	1.604
2971	83.469	20.631
2962	90.604	16.003
2227	7.791	10.702
2958	21.103	21.867
2774	9.903	9.732
2986	81.181	24.179
2207	16.636	26.074
2977	10.734	10.881
2781	17.560	30.725
2235	4.866	26.727
2160	13.263	28.741
2243	30.053	18.465
1654	30.000	26.611
2916	89.276	11.774
1867	40.501	23.434
1897	30.406	27.627

(1 of 18) 1 2 3 4 5 6 7 8 9 10 25

Download Summary per Nanotube:

Figure 15. NanoImage: Output formats available - Results page for Nanotubes

While the previous results (Figure 15) show measurements along each nanotube as one entry, the measurements made in points along the nanotubes are shown in the next table (Figure 16). Each nanotube has its own "ID", as shown in Table 15. Along each nanotube ID, we notice in the first column that a series of measurements has been made ("Pos."). There are measurements available for each nanotube, which are explained below. The NanoImage application makes available the functionality of the Ridge Detection plugin over the web, both as a web application and as a web service (presented in the next section). In that spirit, only a brief explanation of its functionality is provided here. For more information, or to view the source code of the Ridge Detection plugin and questions about its results, please refer to https://github.com/thorstenwagner/ij-ridgedetection.

The results provide numerical values that quantify properties of the CNTs that affect their functionality in a straightforward fashion. Such are Angle of Normal, Line Width and Estimated Length that describe straightness vs curviness, Width and Length respectively. These are more straightforward to use for investigating their relationship with toxicity. Other results, like Contour Class, Contrast, Asymmetry, X coordinates/Y coordinates and the Pos/ID measurement identification values provide insight to characteristics that affect behavior on a second level. For example, the presence and number of identified junctions is an indication of how intertwined the nanotubes appear to be under the microscope.

As some calculated features are necessary to better define the overall results, but do not relate to CNT behaviour, while others provide values more critical to CNT characterisation and behaviour, we make a distinction, grouping the first set as Measurement identification values and the latter as descriptors.

First the measurement identification values are:

- The column **Pos** contains the numbering of each point in the CNT. Its value is assigned in the plugin.

- The column **ID** refers to the numbering of the CNT the point in column Pos belongs to. Its value is assigned in the plugin. It is mentioned as Contour ID when viewing results grouped by CNT, as mentioned above.
- The coordinates of the line points are given in the columns **X coordinates** and **Y coordinates** (named arrays *row* and *col* internally in the plugin).

The descriptors are:

- The column **Angle of Normal** contains the direction of the normal to each line point. This gives a measure of the straightness or curviness of the nanotube.
    - **Contour Class** can assume values that characterise the CNT line (if both end points of the line are junction, only the start point, only the end point or no end point is a junction).
- The columns **Contrast** and **Asymmetry** (*contrast* and *asymmetry* in the plugin) contain the true contrast and asymmetry of each line point if the algorithm was instructed to apply the width and position correction. Otherwise, they are set to NULL.
- The **Line Width** is the sum (WidthL +WidthR), which contain the width information for each line point if the algorithm was requested to extract it; otherwise they are NULL.
- **Estimated Length**

Measurements along nanotubes

(1 of 878) 1 2 3 4 5 6 7 8 9 10 25 Columns

Pos.	ID	Angle of Normal	Countour Class	Contrast	Asymmetry	Line Width	Estimated Length	X coordinates	Y coordinates
1	1898	3.142	no_junc	0.000	0.672	30.689	19.418	651.000	178.552
2	1898	3.156	no_junc	0.000	0.672	30.689	19.418	650.037	178.561
3	1898	3.169	no_junc	0.000	0.672	30.689	19.418	649.070	178.583
4	1898	3.179	no_junc	0.000	0.672	30.689	19.418	648.099	178.621
5	1898	3.186	no_junc	0.000	0.672	30.689	19.418	647.118	178.675
6	1898	3.187	no_junc	0.000	0.672	30.689	19.418	646.126	178.743
7	1898	3.182	no_junc	0.000	0.655	27.003	19.418	645.122	179.026
8	1898	3.168	no_junc	0.000	0.704	28.880	19.418	644.087	179.294
9	1898	3.144	no_junc	0.000	0.742	31.176	19.418	643.007	179.467
10	1898	3.308	no_junc	0.000	0.715	10.421	19.418	643.333	184.911
11	1898	3.288	no_junc	0.000	0.691	31.698	19.418	641.350	179.377
1	1879	3.304	no_junc	0.000	0.376	7.902	17.790	49.252	534.433
2	1879	3.365	no_junc	0.000	0.413	9.086	17.790	48.074	533.929
3	1879	3.412	no_junc	0.000	0.442	10.036	17.790	47.075	533.662
4	1879	3.487	no_junc	0.000	0.479	11.254	17.790	45.718	533.433
5	1879	3.588	no_junc	0.000	0.550	18.873	17.790	44.501	533.867
6	1879	3.588	no_junc	0.000	0.599	24.234	17.790	43.932	534.768
7	1879	3.588	no_junc	0.000	0.591	27.601	17.790	43.273	535.481
8	1879	3.588	no_junc	0.000	0.586	29.347	17.790	42.950	535.807
9	1879	3.588	no_junc	0.000	0.583	32.414	17.790	42.225	536.380
10	1879	3.588	no_junc	0.000	0.590	34.228	17.790	41.439	536.827
11	1879	3.588	no_junc	0.000	0.594	35.118	17.790	41.060	537.035
12	1879	3.588	no_junc	0.000	0.601	37.741	17.790	40.285	537.507
13	1879	3.588	no_junc	0.000	0.729	30.970	17.790	38.953	536.813
14	1879	3.588	no_junc	0.000	0.789	27.440	17.790	38.147	536.128

(1 of 878) 1 2 3 4 5 6 7 8 9 10 25

Download Measurements along nanotubes:

Figure 16. NanoImage: Measurements along each specific nanotube.

Integration of NanoImage into the NanoCommons infrastructure - Availability as a web service

As mentioned before in this deliverable, the NanoImage applications have been registered in the Jaqpot platform and can be accessed through the latter's Descriptor API, https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/. More specifically, they are registered under the ids *'nanoImageSpherical'* and *'nanoImageNanotubes'*.

In order to provide a more thorough description of the functionality of the service, a walk-through of the programmatic access to the services will be laid out next. This is intended mainly for developers who wish to add this functionality to their services by making calls to the NanoCommons infrastructure, or users with some technical proficiency in web services. The User Interface should appeal to most users for everyday purposes.

Under the operation https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/getDescriptor one can request the service description for the Spherical particles with the proper *id* (Figure 17):

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	nanoImageSpherical		path	string

Figure 17. Parameter to GET Spherical particles service description.

And the JSON received in response to this request is as follows:


```
{
 "meta": {
          "descriptions": [
          "Descriptors for Spherical particles"
          ],
          "titles": [
          "NanoImage Spherical"
          ],
          "subjects": [
          "imageS"
          ],
          "creators": [
          "6d9b8261-e255-4fca-9289-0fde48644cf8"
          ],
          "date": "2018-12-18T16:06:39.132+0000",
          "locked": false
 },
 "ontologicalClasses": [
          "ot:Descriptor"
 ],
 "parameters": [
          {
          "name": "Threshold Algorithm",
          "value":  "Default",
          "scope": "OPTIONAL",
"allowedValues":                                   ["Default","Huang","Triangle","Intermodes","IsoData","IJ_IsoData",
          "Li","MaxEntropy","Mean","Moments","Otsu","RenyiEntropy","Shanbhag"],
          "_id": "filter"
          }
 ],
 "ranking": 0,
 "descriptorService": "http://jaqpot-image-service.jaqpot:8080/nanoImage/descriptors/spherical/calculate",
 "_id": "nanoImageSpherical"
}
```


Under the same operation, https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/getDescriptor, we can access the service description for the Nanotubes image analysis (Figure 18):

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	nanoImageNanotubes		path	string

Figure 18. Parameter to GET Carbon Nanotubes service description

The following JSON is received after executing this command:


```
{
 "meta": {
        "descriptions": [
        "Descriptors for Nanotubes"
        ],
        "titles": [
        "NanoImage Nanotubes"
        ],
        "subjects": [
        "imageN"
        ],
        "creators": [
        "6d9b8261-e255-4fca-9289-0fde48644cf8"
        ],
        "date": "2018-12-18T15:53:34.779+0000",
        "locked": false
 },
 "ontologicalClasses": [
        "ot:Descriptor"
 ],
 "parameters": [
        {
        "name": "sigma",
        "value": [
        "8.58"
        ],
        "scope": "OPTIONAL",
        "_id": "sigma"
        },
        {
        "name": "lowt",
        "value": [
          "0.00"
        ],
        "scope": "OPTIONAL",
        "_id": "lowt"
        },
        {
        "name": "uppt",
        "value": [
        "0.17"
        ],
        "scope": "OPTIONAL",
        "_id": "uppt"
        }
 ],
 "ranking": 0,
"descriptorService": "http://jaqpot-image-service.jaqpot:8080/nanoImage/descriptors/nanotubes/calculate",
 "_id": "nanoImageNanotubes"
}
```

The **meta** field contains information regarding the title, description, creator of the descriptor service, date of registration, and some ontological annotations that help with search and other internal functionalities of the Jaqpot platform. The **parameter** field contains information of the required (or optional) parameters of the descriptor services. In *nanoImageSpherical* the parameter is the thresholding filter. In *nanoImageNanotubes* the parameters are the sigma, lower and upper threshold. The **descriptorService** field contains the endpoint where the descriptor service resides.

In order to use descriptor services one can POST on the Descriptor resource, through https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/applydescriptor, a request on a specific descriptor by providing:

1. "title": a title for the produced final descriptors Dataset,
2. "description": a description of the dataset that will be returned as the output of the service,
3. "dataset_uri": the specific dataset Uniform Resource Identifier (URI) from Jaqpot that will be used to produce the new dataset of descriptors,
4. "description_features": an array of strings specifying the features of the dataset that the service takes as an input (in our case a column that contains base64 images),
5. "parameters": the parameters of the service supplied as an array of values and/or strings.

We present next an example of applying the *nanoImageSpherical* service on the image of gold NMs depicted in Figure 5. The image is contained in a csv file in Base64 format that can be found at the following address:

https://github.com/KinkyDesign/descriptor-csv-examples/blob/master/spherical_image.csv.

Before using the service, we need to upload the csv file to the Jaqpot platform, using the procedure described in Annex 1. One can see the dataset by GETting on the Dataset API, https://api.jaqpot.org/jaqpot/swagger/#!/dataset/getDataset, with the id of the dataset, in our case, *317a2a94a30940e2bf2b2702c603b2b2* (Figure 19). The operation can return either application/json or text/csv depending on the 'Response Content Type' field. In addition, the user can select if he wants the data entries or just the meta information of the dataset by checking true or false on the *dataEntries* field.

Response Content Type: application/json ▼

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	317a2a94a30940e2bf2b2702c603b2b2		path	string
dataEntries	true ▼		query	boolean

Figure 19. Parameters to GET a Jaqpot Dataset

The resulting JSON is:


```
{
  "meta": {

    "descriptions": [
      "Spherical Dataset without spherical descriptors"
    ],
    "titles": [
      "Initial Spherical Dataset"
    ],
    "creators": [      "6d9b8261-e255-4fca-9289-0fde48644cf8"
    ],
    "date": "2018-12-29T14:09:27.398+0000",
    "locked": false
  },
  "visible": true,
  "featured": false,
  "dataEntry": [
    {
      "datasetId": "317a2a94a30940e2bf2b2702c603b2b2",
      "entryId": {
        "name": "row1",
        "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
        "URI": "https://api.jaqpot.org/jaqpot/services/substance/862bfbcaa8c644cb9b0b13c983f1103c"
      },
      "values": {
        "https://api.jaqpot.org/jaqpot/services/feature/image_991bc5dcced24a028c5abfe72653929d":
"data:image/jpeg;base64...omitted",
        "https://api.jaqpot.org/jaqpot/services/feature/image_source_eeca1229fc0440619e50de84c99f0057": "10.1021/nn406018q"
      },
      "_id": "f73ba0cbf1234d50ae177e9c82179be8"
    }
  ],
  "features": [
    {
      "name": "image",
      "units": "NA",
      "conditions": {},
      "category": "EXPERIMENTAL",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/image_991bc5dcced24a028c5abfe72653929d"
    },
    {
      "name": "image_source",
      "units": "NA",
      "conditions": {},
      "category": "EXPERIMENTAL",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/image_source_eeca1229fc0440619e50de84c99f0057"
    }
  ],
  "totalRows": 1,
  "totalColumns": 2,
  "_id": "317a2a94a30940e2bf2b2702c603b2b2"
}
```


The application of the POST request on the Jaqpot Descriptor service is shown in Figure 20 with the following parameters:

1. “title”: “Spherical Images Dataset”
2. “description”: “Spherical Images Dataset after applying descriptors service”
3. “dataset_uri”: “https://api.jaqpot.org/jaqpot/services/dataset/317a2a94a30940e2bf2b2702c603b2b2”

4. “description_features”:“https://api.jaqpot.org/jaqpot/services/feature/image_991bc5dcced24a028c5abfe72653929d”
5. {"filter":"Huang"}

The *description_features* URI is retrieved from the *features* field of the dataset and is the URI corresponding to the feature name “image”.

Parameters

Parameter	Value	Description	Parameter Type	Data Type
title	Spherical Images Dataset		formData	string
description	Spherical Images Dataset after applying descrip		formData	string
dataset_uri	https://api.jaqpot.org/jaqpot/services/dataset/317a2		formData	string
description_features	https://api.jaqpot.org/jaqpot/services/feature/image_991bc5dcced24a028c5abfe72653929d		formData	Array[string]
parameters	{"filter":"Huang"}		formData	string
id	nanoImageSpherical		path	string
Authorization			header	string

Figure 20. Parameters to POST on the *nanoImageSpherical* Descriptor service

The POST operation initiates a descriptor calculation procedure. The user can access information about this task through the Task API, https://api.jaqpot.org/jaqpot/swagger/#!/task/getTask, using the id of the task that is returned from the POST operation. When the asynchronous task is finished successfully, the task reports the id of the resulting dataset. In our case, this dataset got the id 6bbf251b298d48b0a1b6c166a7933ffb and, as presented earlier, we can GET it from the Dataset API, https://api.jaqpot.org/jaqpot/swagger/#!/dataset/getDataset (Figure 21).

Response Content Type application/json ▼

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	6bbf251b298d48b0a1b6c166a7933ffb		path	string
dataEntries	true ▼		query	boolean

Figure 21. Parameters to GET the resulting Dataset

The JSON received is presented next:

{
  "meta": {
    "comments": [
      "Created by task TSK2ff36d1ebefd4dd7a35f8cdce0d4c5c7"
    ],
    "descriptions": [
      "Spherical Images Dataset after applying descriptors service"
    ],
    "titles": [
      "Spherical Images Dataset"
    ],
    "creators": [
      "6d9b8261-e255-4fca-9289-0fde48644cf8"
    ],
    "hasSources": [
      "https://api.jaqpot.org/jaqpot/services/dataset/317a2a94a30940e2bf2b2702c603b2b2"
    ],
    "date": "2018-12-29T19:27:36.543+0000",
    "locked": false
  },
  "visible": true,
  "dataEntry": [
    {
      "datasetId": "6bbf251b298d48b0a1b6c166a7933ffb",
      "entryId": {
        "name": "Spherical image: 0"
      },
      "values": {
        "https://api.jaqpot.org/jaqpot/services/feature/0a79eba090e74327bbf4cc562738a71a": 268.83417,
        "https://api.jaqpot.org/jaqpot/services/feature/0cc9f57e366b44969da0e1d4038ef533": 0.007945704353583844,
        "https://api.jaqpot.org/jaqpot/services/feature/1708e57ef4f64a33982cf00de4ec352e": "data:image/jpeg;base64...omitted",
        "https://api.jaqpot.org/jaqpot/services/feature/270cec01d7244923b22146589ba6d28c": 121.94643,
        "https://api.jaqpot.org/jaqpot/services/feature/33ffb74c1e8c4c978ea91ed4f275b4bc": 31696.383733607567,
        "https://api.jaqpot.org/jaqpot/services/feature/350659a9f806443592ebdeea46b759f2": 21.227932,
        "https://api.jaqpot.org/jaqpot/services/feature/4b1f089be9aa4824976a1864f0c90797": 207.7183,
        "https://api.jaqpot.org/jaqpot/services/feature/4cd07c44d5934f9cbf9fe05a5f136e29": 755.125,
        "https://api.jaqpot.org/jaqpot/services/feature/5e4575f1e6d14da4b26d69e4e02eb773": 0.6408267,
        "https://api.jaqpot.org/jaqpot/services/feature/5ffcadb85bc9476b840fa352fc517a23": 207.7183,
        "https://api.jaqpot.org/jaqpot/services/feature/61b3fed0a77c4df298f44a7bb456241c": 207.7183,
        "https://api.jaqpot.org/jaqpot/services/feature/6687f9feb7784428b0091a736a290222": 36.410713,
        "https://api.jaqpot.org/jaqpot/services/feature/6e664a4819cf45289d64361c4b7b9750": 207.74889,

...

      },
      "_id": "bd922316c1bf4873952b452bb255c600"
    }
  ],
  "features": [
    {
      "name": "FeretAngle",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/6ea0a812f1b84bd29485b908ed6c2a82"
    },
    {
      "name": "Raw Integrated Density",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/9bf75f2b8fa24e758d454346edbef745"
    },
    {
      "name": "Minor",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/61b3fed0a77c4df298f44a7bb456241c"
    },

....

    }
  ],
  "totalRows": 1,

"totalColumns": 34,
"existence": "DESCRIPTORSADDED",
"_id": "6bbf251b298d48b0a1b6c166a7933ffb"
}

Apart from the column containing the image, the descriptor services also accept a column with the scaling factor (pixel to nm ratio). The scaling factor is taken into consideration in order to return results in the correct ratio.

NanoXtract: Nanoparticles Image Analysis Tool Powered by Enalos Cloud Platform

Introduction

The Enalos Cloud platform, developed by NovaMechanics Ltd (http://www.novamechanics.com/), is an online, freely available toxicity and drug discovery platform, that hosts predictive models released as web services, which aim to address the need to reduce the amount of time and cost spent on experimental testing during the drug discovery and the risk assessment procedures for small molecules and NMs (Varsou et al., 2019 and Melagraki & Afantitis, 2014).

One of the platform's recent features is to analyze TEM microscopy images depicting nanoparticles (NPs) and extract useful descriptors that can be used in a future step as inputs to toxicity prediction models. The image analysis tool allows the user to upload an image and in a short time with simple steps it is possible to achieve satisfactory labeling of the NPs. The extracted geometric indices can be used in predictive models already existing in the Enalos Cloud platform or in in-house models.

The tool can be easily accessed through the link:

http://enaloscloud.novamechanics.com/EnalosWebApps/NanoXtract/

Application workflow

Step 1: Image upload

To initiate the image analysis, the user must upload a TEM image in .png or .tiff format (maximum size 5MB). By clicking on the **Upload** button (Figure 22A) a popup file upload window appears, and the user can browse and select the photo of interest. It is strongly advisable to import high quality images in order to achieve an effective segmentation of the image. Some chara

In the next step, the user must select from the **NP type** dropdown menu the type (circular/cylindrical/plates/other) of the depicted particles in the uploaded image (Figure 22B). The user must also select the computational methods that will be applied for the calculation of the image descriptors: with and without outliers / with outliers only (Figure 23B). In the default settings "with and without outliers" is selected, and in later steps an additional statistical filtering is activated (Mondini et al., 2012). The detailed user guide along with some characteristic examples can be found on the top ribbon (Figure 23A).

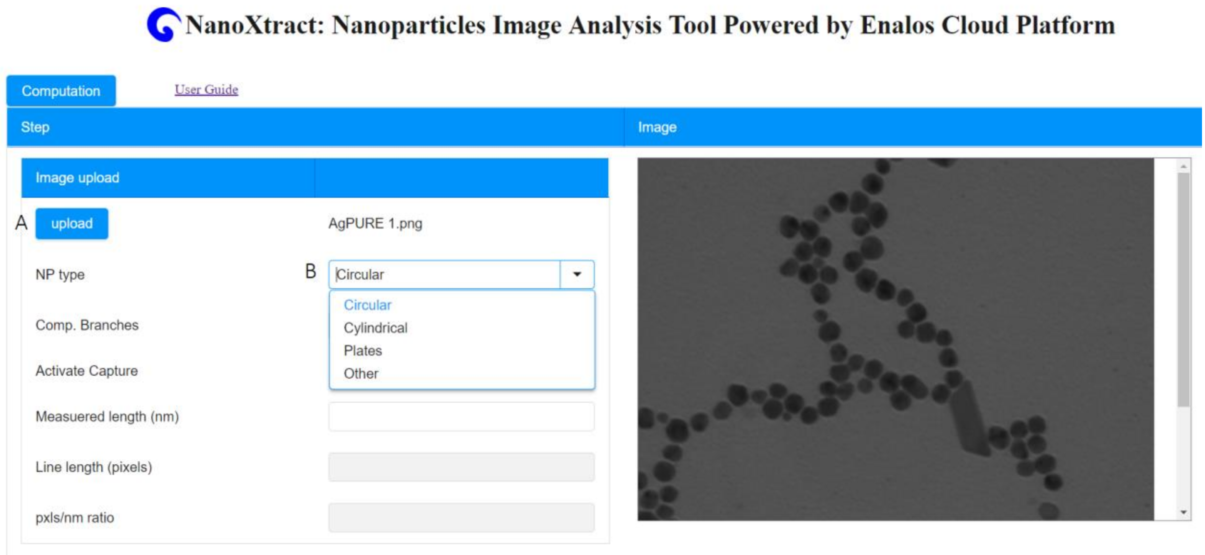


Figure 22: Image upload parameters. [A] Browse button. [B] Nanoparticle type list: Circular/Cylindrical/ Plates/Other.

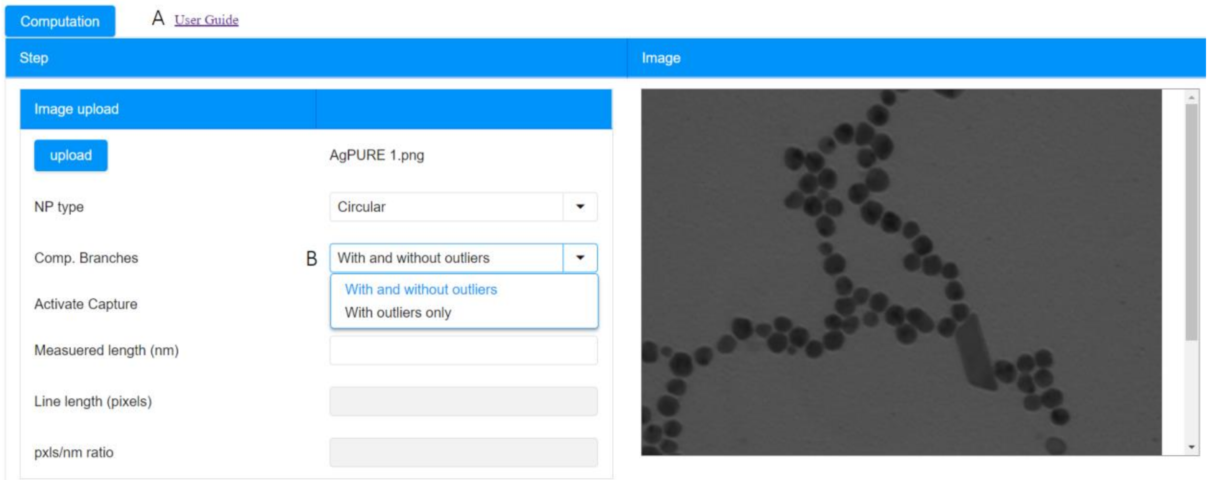


Figure 23: Filtering of the statistical outliers.

By clicking on the **Activate Capture** checkbox (Figure 24A) the user can draw a line and measure the scale bar (ruler) in pixels (Figure 24C). Then the user should type the known distance of the ruler in nm, in the **Measured Length** field (Figure 24B). In the **Line Length** field, the measurement of the line in pixels will be displayed whereas, and the scale ratio will be displayed in the **pxls/nm ratio** field

(Figure 24D). This step is important in order that the calculated descriptors, such as the diameter, aer displayed in the proper units. After the scaling measurement, the user should uncheck the **Activate Capture** box, in order to proceed with the further steps of the analysis.

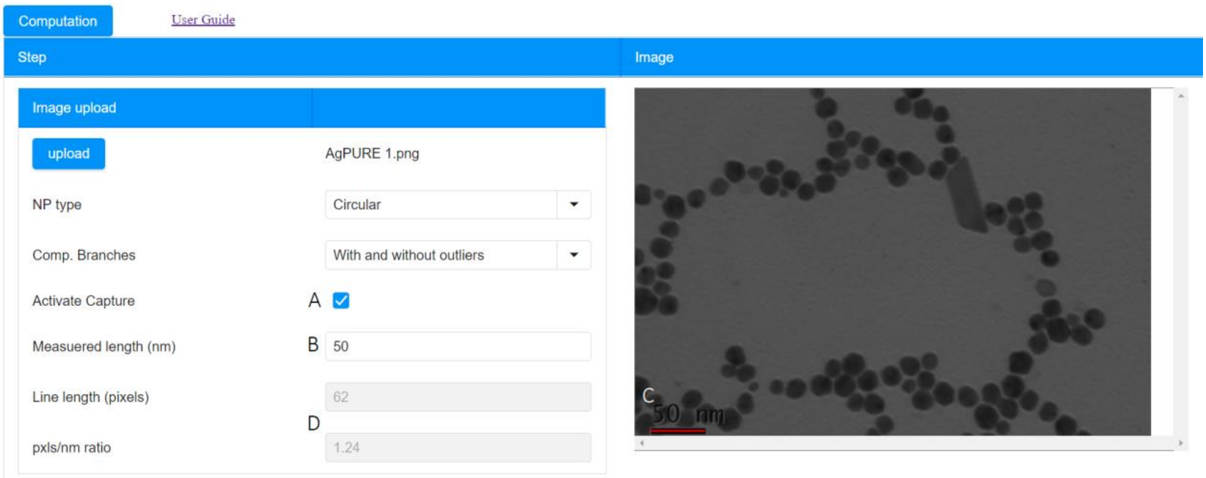


Figure 24: [A] Activate Capture button enables the line measurement. [B] The user should type the scale value depicted on the image scale bar in the appropriate field. [C] The measured line length of the scale bar in pixels is displayed automatically, as well as the pxls/nm ratio [D].

After the initial steps the user should tune a series of parameters in order to achieve a satisfactory segmentation and labelling of the NPs (left side of the interface). In every step of the image analysis, the user can inspect the results of the tuning in order to adjust the parameters in the most efficient way (right side of the interface). By right-clicking on the images, the user can save them for later use. The effects of the tuning are displayed every time the user clicks on the **Computation** button on the top ribbon of the User Interface (UI) (Figure 25A).

Step 2: Reduction of noise

The noise elimination (especially when the analyzed image is of low quality) is an important step, in order to avoid false positives during segmentation (background details may be considered as possible NPs). This step applies a Gaussian blur in order to reduce image detail, the **X, Y dimension sigma** (Figure 25B) values determine the effect of the blurring (higher values lead to more blurring).

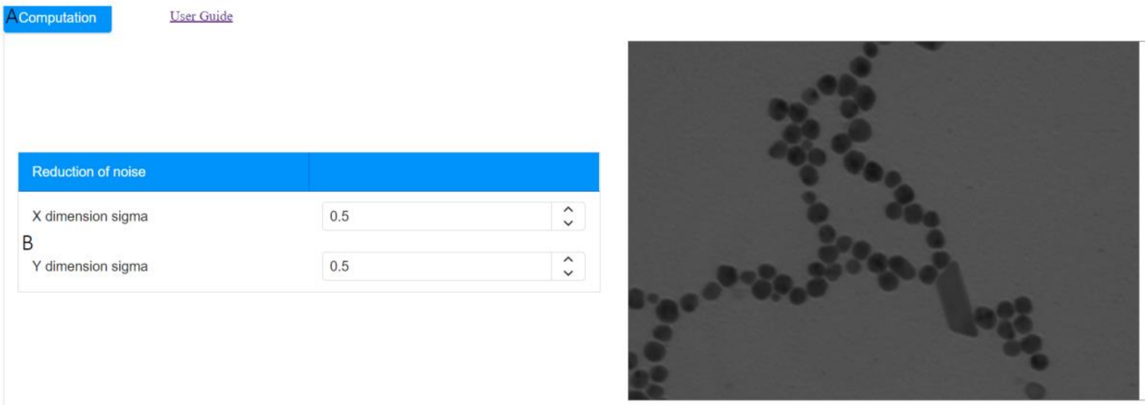


Figure 25: Elimination of noise step. [A] Computation button. In order that the changes of the parameters take place, the user must click every time on this button. [B] The user can change the X, Y dimension values by clicking on the arrows or by directly typing into the appropriate fields.

Step 3: Thresholding

The following step is a thresholding procedure. This technique is commonly used when detecting edges, counting particles or measuring areas of images. The grayscale TEM image is converted to a binary one (only black and white colours) by defining a grayscale cut off limit. Grayscale values below the threshold are converted to black and those above become white. The user can select an automatic threshold method by clicking on the dropdown menu of the **Method** field (Figure 26). In the case that the Manual method is selected (Figure 27A), the user should type a **Threshold Value** into the corresponding field (Figure 27B). The user can observe the effects of the thresholding on the right side of the screen.

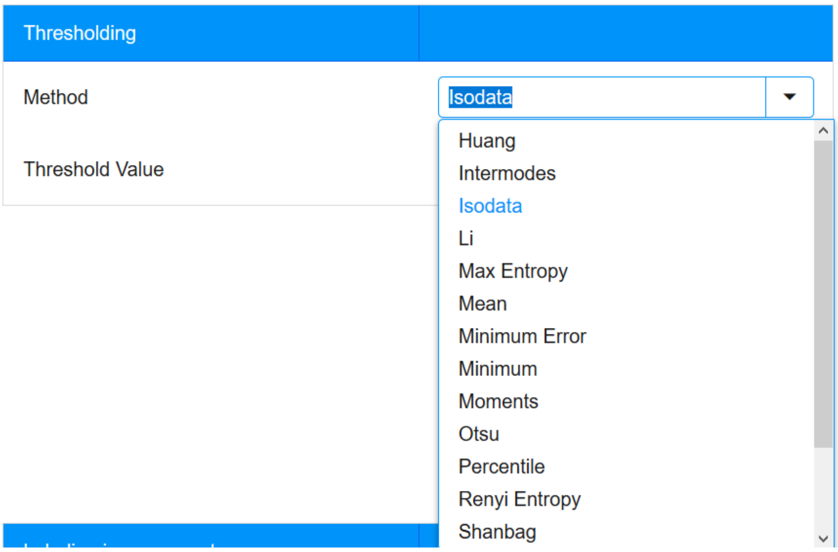


Figure 26: Thresholding methods menu.

Figure 27: [A] Manual thresholding selected. [B] Threshold value. [C] Aggregated NPs.

Step 4: Labeling improvement

In case of unsatisfactory segmentation (unseparated segments), the user can improve the quality of the extracted labeling (Figure 28). **Distance Merge Threshold** controls the proximity (in pixels) of the segments that will be merged, **Gauss Size** controls the extent of the splitting (higher values lead to less splitting), whereas **Size Merge Threshold** controls the minimum size (in pixels) of segments that will be merged regardless of their shape, presuming there is another object close enough.

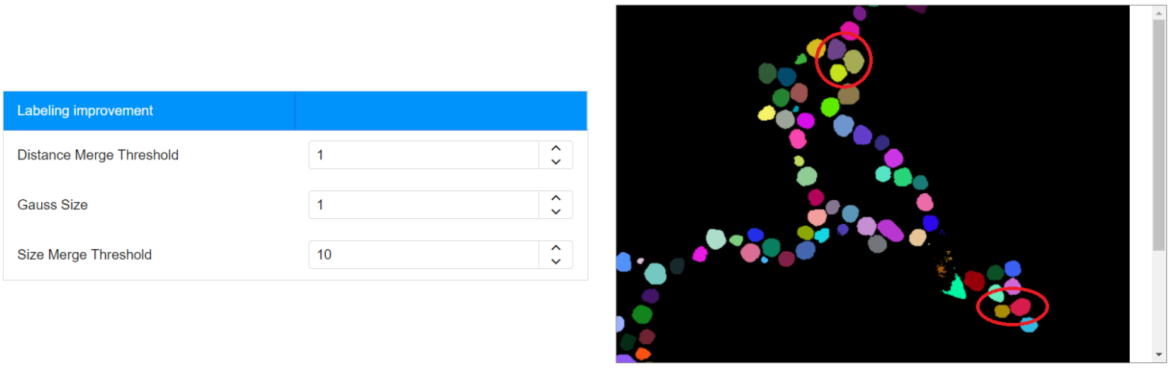


Figure 28: Labeling improvement field. The combination of the field parameters leads to segment separation.

Step 5: Filtering

When satisfactory segmentation is achieved, the labels are filtered based on their size. The user can adjust the filtering levels (**Minimum Segment Area** and **Maximum Segment Area** pixel-size cut-offs) in order to remove outlier labels (Figure 29): small labels may possibly encode noisy parts of the image, whereas bigger labels may represent aggregates that could not be split. Segments touching the border of the image are also removed as these will distort the mean particle size and other descriptor averages.

Figure 29: Filtering step. Removal of segments touching the border as well as removal of segments with less pixels than the minimum segment threshold and more pixels than the maximum segment area threshold.

After the above preparatory steps are completed, the image descriptors (mean values of all segments) are ready to be calculated, by clicking on the **Computation** button. In this step the segments are filtered according to their shape type as selected at the beginning (Figure 30).

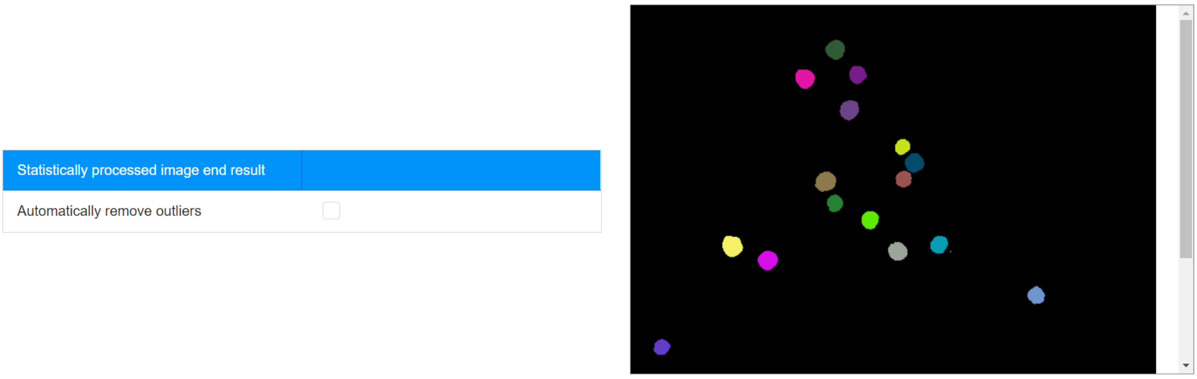


Figure 30: Filtered segments according to the tuned parameters and the shape of the particles.

By clicking on the **Automatically remove outliers** box (Figure 31) the results of the statistical filtering results are displayed. The statistical filtering is performed in order to exclude the outliers of each descriptor and re-calculate the mean descriptor values. In detail, z-scores are calculated for the descriptors of main elongation, circularity, eccentricity and solidity. Later the segments with a score exceeding the percentile 0.65 based on the z-score, are excluded (Mondini et al., 2012).

The aim of this step is to remove any segments that have significant differences from most segments in the image. Differences may occur due to “false-positive segments” that it wasn’t possible to filter in the previous steps (e.g. noise, aggregates, segmented scale bar, particles with different shapes or size etc.). Nevertheless, in many cases this variability in shape and size may add crucial information to the analysis and for this reason the user can select to download the calculated descriptors both before and after the statistical filtering, by checking/un-checking the **Automatically remove outliers** box (Figure 31) and by clicking on the Descriptors download buttons (Figure 32).

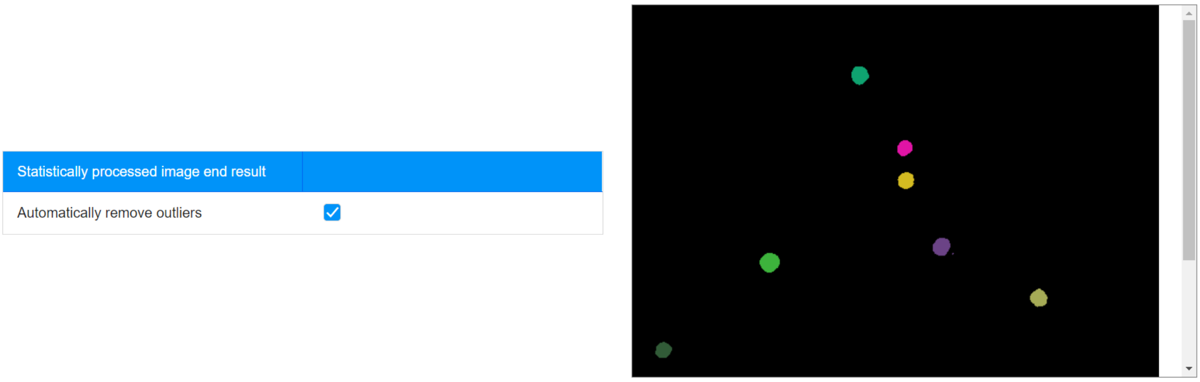


Figure 31: Filtering of the statistical outliers.

Step 6: Results

For each segment of the final stage, 16 image descriptors are calculated (Table 3). Then considering all segments, the mean value of each descriptor is calculated and presented in a table (Figure 32). By clicking on **Descriptors mean** values button (Figure 32C), the user can download a .csv file containing the mean descriptor values and their standard deviation (Figure 33) and by clicking on **Descriptors per component** button (Figure 32D), the user can download a .csv file containing the calculated descriptors for each particle (Figure 34).

Table 3: Calculated image descriptors

Image descriptor	Brief meaning	Range of values
Boundary size	Total length of the NP's boundary (the perimeter calculated by different method)	>0 [nm]
Boxivity	The extent to which a NP approaches a rectangle	0-1 [unitless]
Circularity	The degree to which a NP approaches a perfect circle	0-1 [unitless]
Convexity	The NP's edge roughness	0-1 [unitless]
Diameter	The NP's diameter	>0 [nm]
Eccentricity	The measure of how much the NP deviates from being circular	0-1 [unitless]
Extent	The boxivity calculated using a different method	0-1 [unitless]

Main elongation	The lengthening of the NP	0-1 [unitless]
Major axis	The longest diameter of the best fitting ellipse to the NP	>0 [nm]
Maximum Feret's diameter	The longest distance between any two points along the selection boundary (caliper diameter)	>0 [nm]
Minimum Feret's diameter	The shortest distance between any two points along the selection boundary	>0 [nm]
Minor axis	The shortest diameter of the best fitting ellipse to the NP	>0 [nm]
Perimeter	Total length of the NP's boundary	>0 [nm]
Roundness	Compares the surface of the NP to the surface of the disc of diameter equal to major axis	0-1 [unitless]
Size	The area of the NP	>0 [nm^2]
Solidity	The degree of the overall concavity or convexity of a NP	0-1 [unitless]

Description	A Values	B Standard Deviation
Circularity	0.814097	0.035058
Perimeter [nm]	59.011633	7.619581
Convexity	0.968588	0.011238
Extend	0.751080	0.027440
Diameter [nm]	19.317530	2.470115
Size [nm2]	209.375581	53.695768
Circularity #2	0.864631	0.018847
Convexity #2	0.961173	0.008305
Eccentricity	0.395465	0.078838
Main Elongation	0.039860	0.029660
Minimum Ferets Diameter [nm]	15.629240	2.148347
Maximum Ferets Diameter [nm]	17.679132	2.430156
Major Axis [nm]	16.922178	2.329364
Minor Axis [nm]	15.476365	2.160625
Boundary Size [nm]	54.668489	7.513076

C Descriptors mean values

D Descriptors per component

Figure 32: Mean values [A] and standard deviation [B] of the calculated image descriptors. By clicking on the buttons [C]-[D] the user can download a .csv file containing the mean descriptor values and their standard deviation and all the computed descriptors per particle respectively.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	row ID	Circularity	Perimeter	Convexity	Extend	Diameter [	Size [nm2]	Circularity	Convexity	Eccentricit	Main Elon	Minimum I	Maximum	Major Axis	Minor Axis	Boundary :	Boxivity	Roundness	Solidity
2	Mean	0.814097	59.01163	0.968588	0.75108	19.31753	209.3756	0.864631	0.961173	0.395465	0.03986	15.62924	17.67913	16.92218	15.47637	54.66849	0.777652	0.914713	0.966768
3	Std. deviat	0.035058	7.619581	0.011238	0.02744	2.470115	53.69577	0.018847	0.008305	0.078838	0.02966	2.148347	2.430156	2.329364	2.160625	7.513076	0.017293	0.031128	0.009388

Figure 33: File containing the mean descriptor values and their standard deviation.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	row ID	Circularity	Perimeter	Convexity	Extend	Diameter [	Size [nm2]	Circularity	Convexity	Eccentricit	Main Elon	Minimum I	Maximum	Major Axis	Minor Axis	Boundary :	Boxivity	Roundness	Solidity
2	Row0	0.787022	66.88359	0.960829	0.755435	21.65548	257.3233	0.834554	0.948899	0.426576	0.043478	17.50488	20.24441	19.03277	17.21422	62.24682	0.756917	0.904452	0.956305
3	Row1	0.820078	65.44313	0.969697	0.753623	20.75244	257.6592	0.875704	0.964843	0.364667	0.090909	17.38787	19.34496	18.77029	17.47773	60.80637	0.774747	0.931138	0.969659
4	Row2	0.822386	57.04752	0.984472	0.754762	18.49265	194.1682	0.888274	0.972889	0.334079	0	15.52535	16.89797	16.19545	15.26494	52.41076	0.792867	0.942545	0.978003
5	Row3	0.831412	62.44571	0.979592	0.793388	20.55727	236.8315	0.890552	0.972991	0.389868	0	16.80828	19.06509	18.0957	16.6638	57.80894	0.79932	0.920871	0.979167
6	Row4	0.839122	72.67955	0.966851	0.747863	23.89734	327.8689	0.860346	0.954768	0.413037	0.036364	19.7947	22.38771	21.40987	19.49828	69.20197	0.762645	0.910714	0.963475
7	Row5	0.807644	63.6049	0.962687	0.764822	20.54092	238.5112	0.861953	0.95809	0.457732	0.045455	16.80828	19.20553	18.48147	16.43169	58.96814	0.768398	0.88909	0.964674
8	Row6	0.820563	52.60964	0.967626	0.707895	17.52259	163.2626	0.849897	0.960465	0.45449	2.22E-15	13.8675	15.91515	15.27636	13.60745	49.13207	0.7776	0.890752	0.960474
9	Row7	0.81556	40.05742	0.962733	0.738095	13.3433	90.70142	0.851805	0.962	0.359928	0.064516	10.53638	11.73592	11.12566	10.38002	36.57985	0.750834	0.93298	0.954064
10	Row8	0.821838	62.6446	0.959799	0.722117	19.94348	235.4878	0.845315	0.951079	0.263472	0.037975	17.27755	18.32843	17.62993	17.00701	59.16702	0.758911	0.964667	0.95896
11	Row9	0.823235	49.61222	0.97561	0.784314	16.5971	144.7863	0.899469	0.978074	0.495373	0.058824	12.75111	15.06949	14.56763	12.65461	44.97546	0.792279	0.86868	0.979545
12	Row10	0.850721	61.00525	0.979112	0.741107	20.09451	231.4566	0.878872	0.967438	0.157332	0.033333	16.80828	18.49265	17.27473	17.05958	57.52768	0.791954	0.987546	0.980085
13	Row11	0.848317	52.41076	0.961672	0.764543	17.291	167.2937	0.837811	0.947607	0.323289	0	14.48989	15.91515	15.00308	14.19742	50.09238	0.768519	0.9463	0.955854
14	Row12	0.814004	65.92328	0.972158	0.759058	21.77923	259.3389	0.867656	0.962404	0.485806	0.043478	17.35186	20.24441	19.43643	16.98874	61.28652	0.762846	0.874067	0.966208
15	Row13	0.838738	53.0898	0.972222	0.736842	17.69429	170.3171	0.869542	0.963607	0.404872	0.038462	14.48989	16.104	15.40006	14.0814	49.61222	0.78	0.914373	0.963878
16	Row14	0.81892	57.52768	0.975684	0.764286	18.41984	196.5198	0.882783	0.966681	0.397403	0.035714	15.39588	16.89797	16.51281	15.15289	52.89091	0.77381	0.917644	0.971761
17	Row15	0.669906	63.6049	0.93586	0.66875	22.07035	194.1682	0.878817	0.967597	0.478211	0	14.7541	17.42647	16.77785	14.73506	52.69203	0.800554	0.878245	0.973064
18	Row16	0.838655	66.40344	0.973333	0.760417	21.77923	271.4324	0.861417	0.959955	0.42621	0	17.80849	20.24441	19.54573	17.68154	62.92586	0.763705	0.904624	0.969988
19	Row17	0.80332	61.76667	0.970588	0.785714	19.96032	222.7224	0.857525	0.952733	0.380892	0.047619	16.22868	18.41984	17.5128	16.19267	57.12991	0.789286	0.924619	0.967883
20	Row18	0.826141	62.6446	0.974619	0.758893	20.09451	237.1674	0.88571	0.964533	0.348284	0.064516	16.76948	18.49265	17.94827	16.82451	58.00783	0.785317	0.937389	0.975138
21	Row19	0.814018	55.40818	0.957929	0.704762	18.19967	180.3951	0.840597	0.953123	0.343448	0.074074	14.48989	16.57684	15.63849	14.68723	51.93061	0.795556	0.939172	0.957219
22	Row20	0.819681	51.45045	0.988462	0.751462	17.36854	155.8721	0.893784	0.977593	0.425997	0.08	13.3307	15.8305	14.81088	13.39977	46.81369	0.806957	0.904725	0.978903
23	Row21	0.829039	60.12733	0.96206	0.768398	19.48338	217.6834	0.852392	0.953597	0.346053	0.047059	16.07059	18.05141	17.18765	16.12571	56.64975	0.8	0.938215	0.96
24	Row22	0.805146	63.12475	0.976864	0.785124	20.4918	234.144	0.860121	0.957084	0.499352	0	16.22868	19.20553	18.54973	16.07147	58.48798	0.790249	0.866399	0.970752
25	Row23	0.759382	42.17692	0.946746	0.761905	13.3433	93.38887	0.832747	0.956524	0.465452	0.071429	10.64551	11.82148	11.59079	10.2587	37.54016	0.763736	0.885073	0.942373
26	Row24	0.836898	62.6446	0.970075	0.736742	20.63882	240.1908	0.862198	0.9561	0.34962	0.032258	17.1074	19.01215	18.06711	16.92692	59.16702	0.768817	0.936891	0.967524
27	Row25	0.804768	61.96556	0.976	0.757764	20.24441	225.0739	0.860576	0.959824	0.49065	0.090909	16.12901	18.72731	18.13509	15.80213	57.32879	0.743123	0.871357	0.971014

Figure 34: File containing the calculated descriptors values per particle.

An explanation of the above-mentioned terms, their units, their ontological annotation and their correspondence to the descriptors calculated by NanoImage are provided in Appendix 2.

Case Studies

Examples of different NP types

Apart from circular NPs the NanoXtract tool can process cylindrical particles (nanotubes) and plates. In the examples below the analysis of nanorods (Figure 35) and nanoplates (Figure 36) is presented.

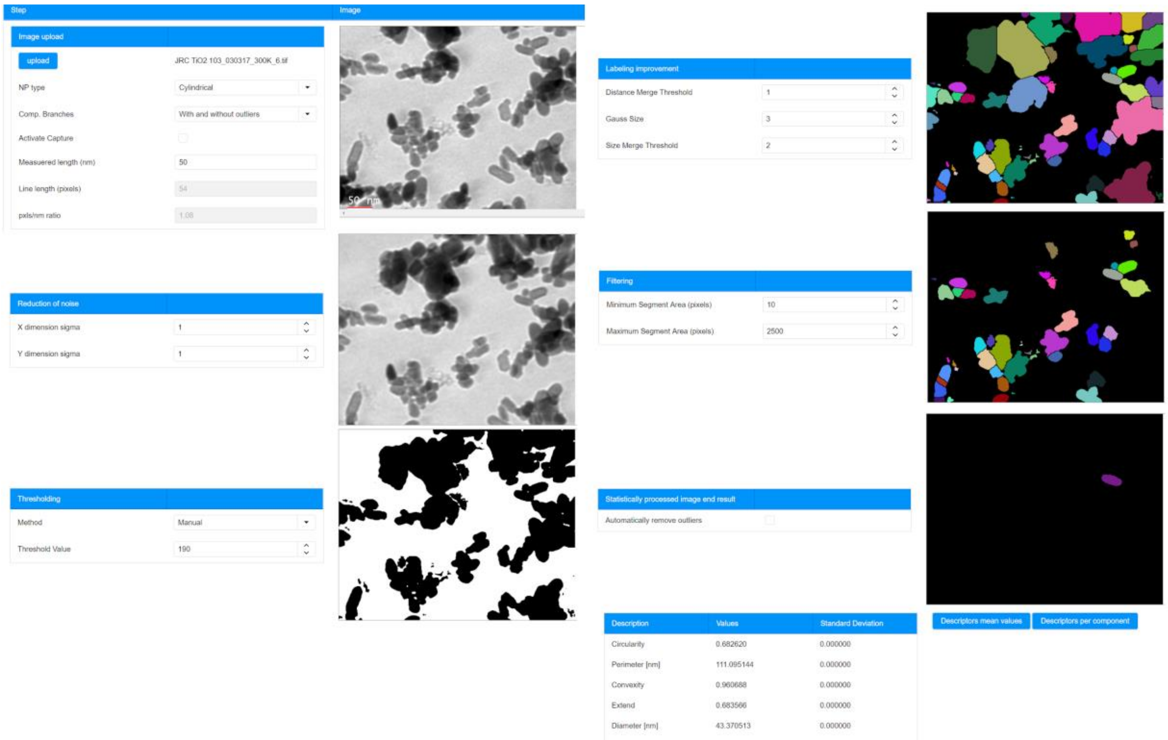


Figure 35: Application of the NanoXtract tool to extract the image nanodescriptors of NPs which have the structure of nanorods.

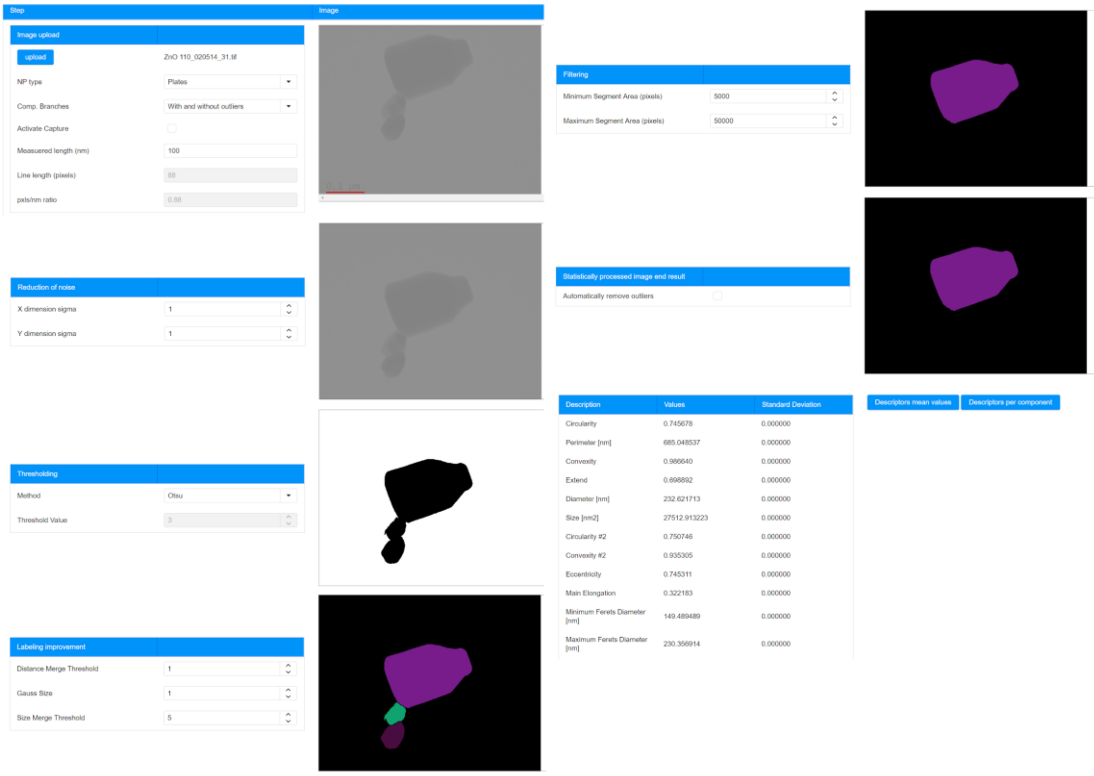


Figure 36: Application of the NanoXtract tool to extract the image nanodescriptors to NPs which have the structure of plates.

If the particles depicted on a TEM image do not fall into the previous categories, the user can select the category “Other” from the dropdown menu of the Image upload field. In Figure 37, the analysis of particles with complex morphology is presented.

This is a characteristic example of sea-urchin shaped particles that result from NM ageing in the environment. CeO_2 NMs (uncoated, provided by Promethean particles to the FP7 NanoMILE project) were artificially aged in phosphate buffer. The protocol was based on trials and modifications of a published method (Zhang et al., 2012). The particles were added to a solution of 5 mM KH_2PO_4, citric acid and ascorbic acid, 5 times the concentration used in the original publication to induce a faster and more complete transformation of the CeO_2 NMs into $CePO_4$ by providing a molar excess of phosphate over Ce. The final concentrations of CeO_2 NMs in the suspensions were 496 mg/L (used for TEM, UV-vis and zeta potential), with analysis performed after 7 and 21 days of static incubation. TEM images were acquired with a JEOL 1200X TEM.

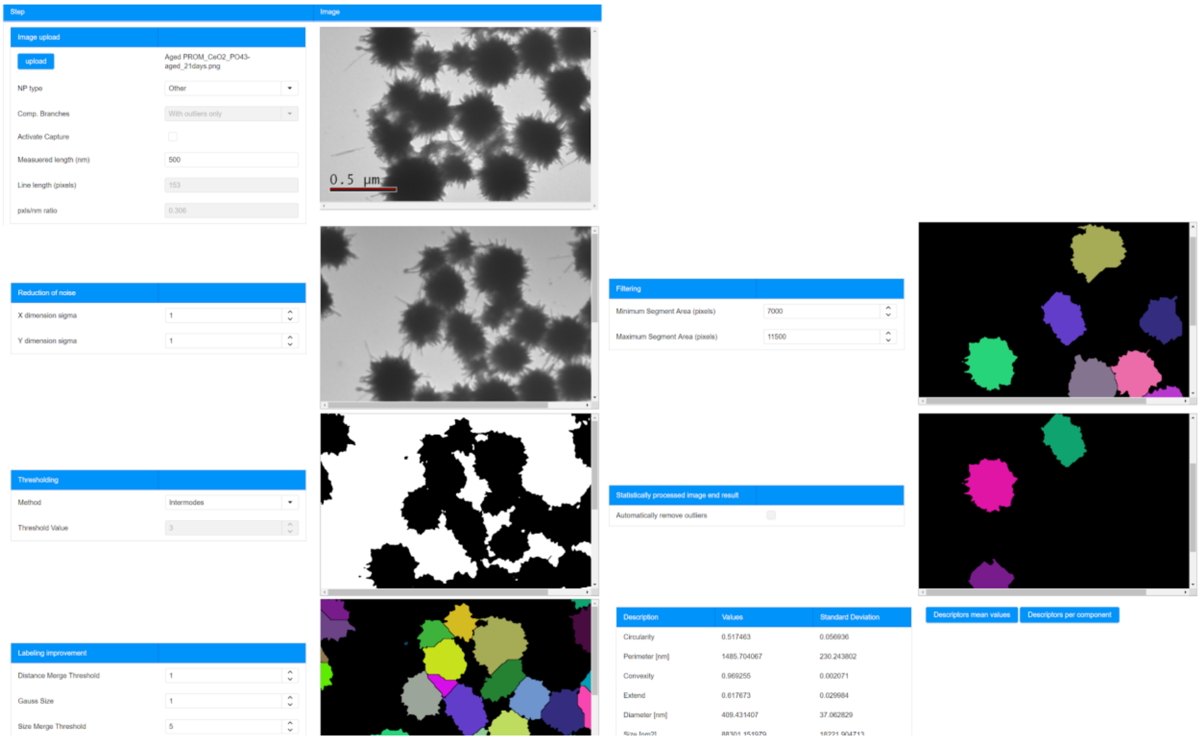


Figure 37: Application of the NanoXtract tool to extract the image nanodescriptors for aged NPs which have complex structures.

Possible errors and treatment

In case that the user selects enough strict parameters that lead to the exclusion of all particles from the analysis, an error message will appear (Figure 38) indicating the inability of the tool to further calculate the particle descriptors. In such a case, the user may try to lessen some of the parameters of the available steps (Figure 39) or try to use another **NP type** from the initial dropdown list, that may be more adequate for successful segmentation and calculation of the descriptors (Figure 40).

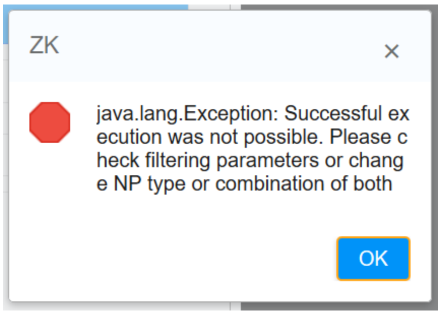


Figure 38: Error message that may be presented in case of strict filtering parameters.

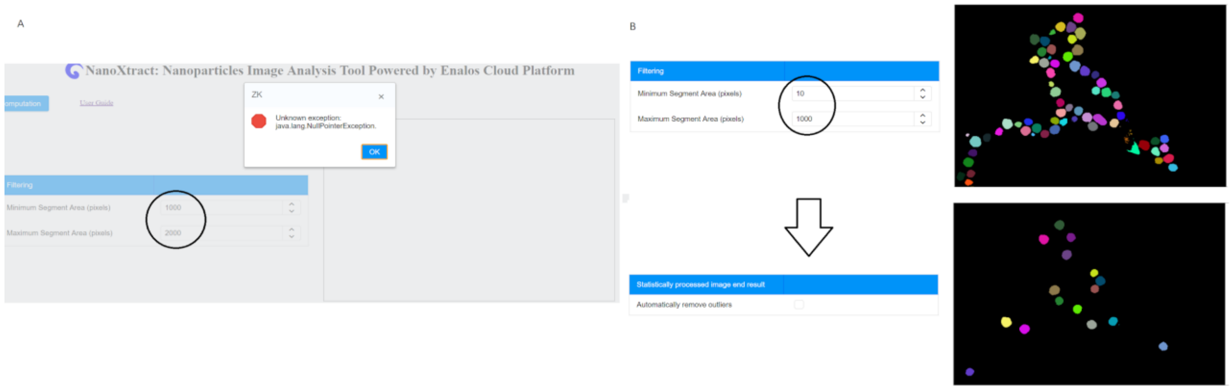


Figure 39: Example of an error produced due to overly strict filtering parameters. [A] Keeping all other parameters unchanged, by slightly altering the Minimum and Maximum Segment Area, the tool is executed successfully, as seen in [B].

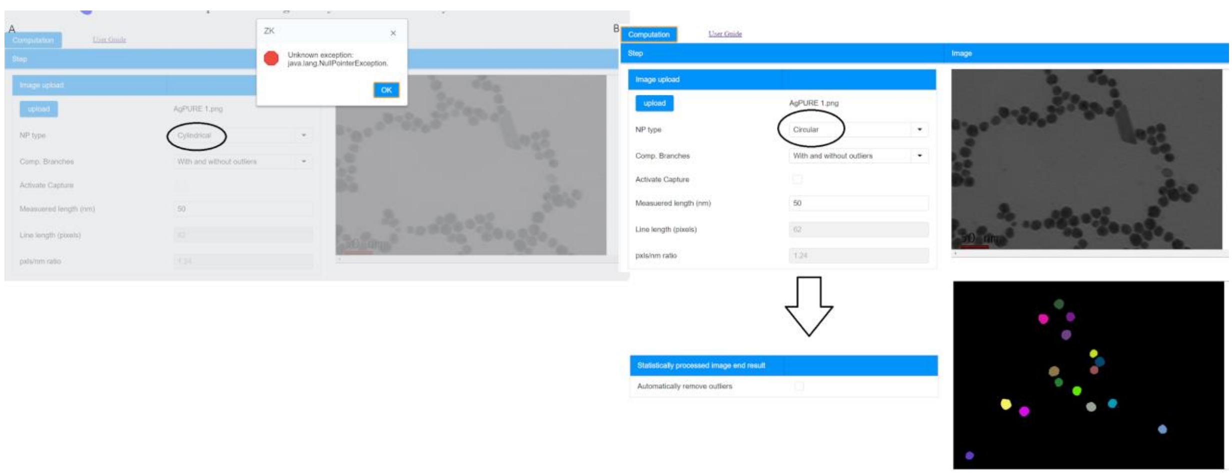


Figure 40: Example of an error produced due to NP type selection. [A] Keeping all other parameters unchanged, by selecting type "Plates" instead of type "Circular", the tool is executed successfully, as seen in [B].

Another possible error may occur when the additional statistical filtering excludes all particles as outliers (even if the **Automatically remove outliers** box is not activated, the statistical filtering is performed). In this case all workflow won't produce any results and an error message will be displayed (Figure 41). The user can select to skip statistical filtering by selecting to present all possible outliers in the final output from the **Comp. Branches** dropdown menu.

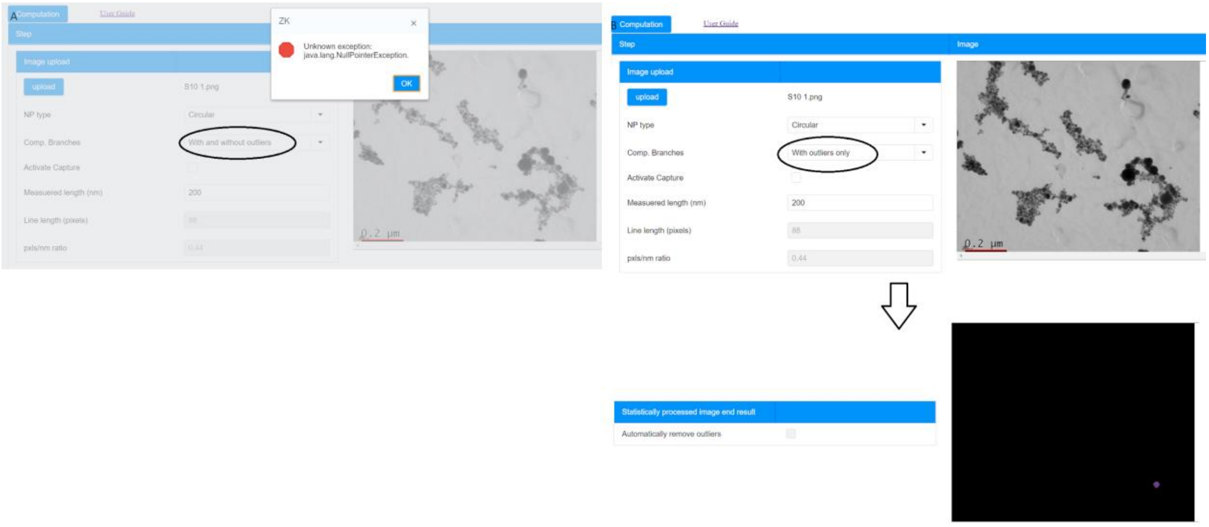


Figure 41: Example of an error produced due to compound branch selection. [A] Both types of filtering will be performed. [B] Selection of only one type of filtering (based on NP shape) leads to the successful execution of the tool.

Suitable images for analysis

Image specifications

A successful segmentation in the image analysis process is greatly affected by the quality of analysed images. There is no concrete rule about the image specs, however there are some simple guidelines that can be followed regarding the image resolution. We have to denote here that the combination of dimensions, resolution and zoom (scale) are responsible for the suitability of an image to be analysed by the presented tool. In Table 4 some examples of suitable images for analysis are presented.

Table 4: Guidelines for the specifications of the analyzed images.

Dimensions [pxls x pxls]	Resolution [ppi]	Zoom [nm]
1024x1024	72	100
1500 x 1000	95	50
770 x 510	220	200

480 x 320	220	100
512 x 512	75	200

Luminosity difference

Even if the uploaded images are of high resolution, luminosity differences in an image may result to unsatisfactory segmentation. In Figure 42A the one half of the TEM image is more luminous than the other half. This leads to unsatisfactory thresholding (step 3) which is a key step for the image analysis process, and thus some NPs are not suitable for calculation of descriptors even if the next steps are applied. TEM image in Figure 42B has a uniform luminosity therefore when step 3 is applied, thresholding (with the same parameters) is better than the previous one.

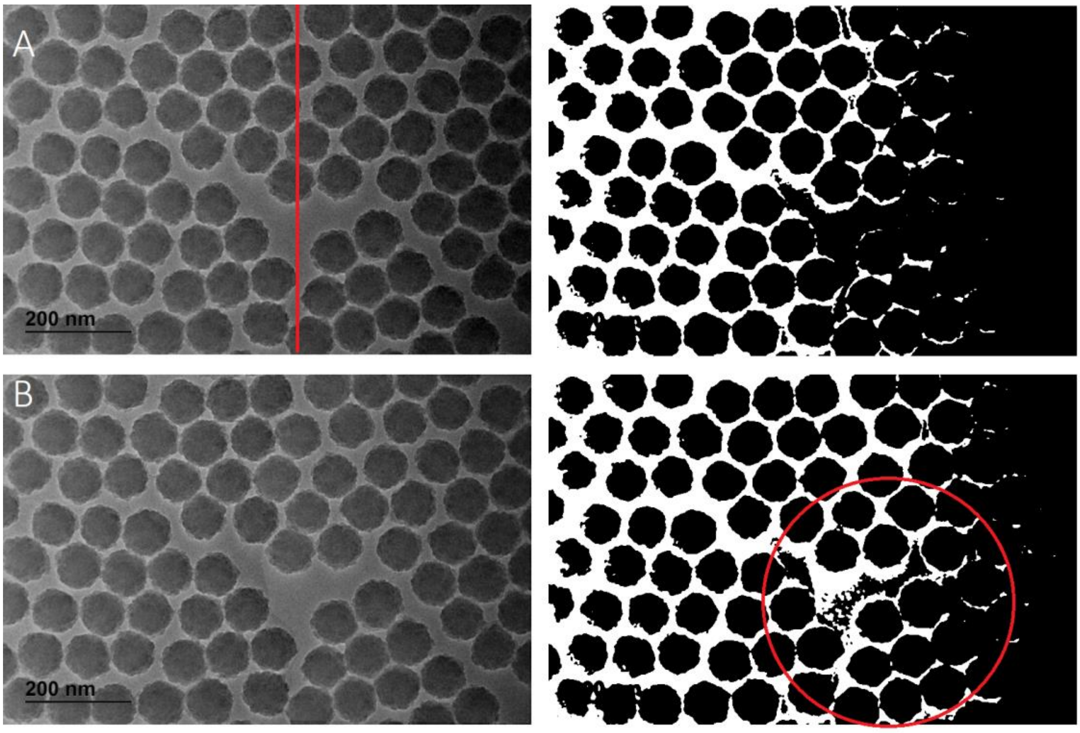


Figure 42: [A] TEM image with non-uniform luminosity and produced thresholding. [B] TEM image with uniform luminosity and improved thresholding.

Aggregation

Aggregated particles can be separated in step 4, “Labelling improvement” as it can be seen in Figure 43B-C (green circles). However, a high aggregation of NPs might be impossible to be separated as shown in Figure 43A and C (red circles). For this reason, it is advised, when higher degree of aggregation is present, to provide more TEM images of the same sample.

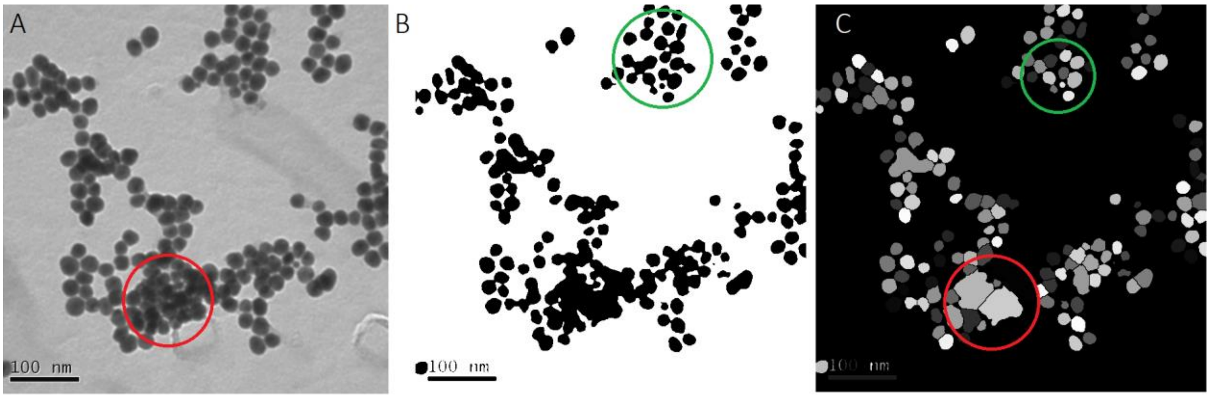


Figure 43: Different degrees of aggregation phenomena. Lower degree of aggregation can be overcome in step 4 (green circles), however highly aggregated particles might be impossible to be separated (red circles).

NanoXtract SOAP web service

The interested users, apart from the above online graphical interface, are welcome to use the implemented NanoXtract SOAP (Simple Object Access Protocol**)** web service. The corresponding WSDL file is located at:

http://enaloscloud.novamechanics.com/EnalosWebServices/wsdl/NanoXtractWebService.wsdl.

Our web service contains a method called "callNanoXtract". Below the input of the method can be seen in xml format:

```
<element name="callNanoXtract">
 <complexType>
  <sequence>
   <element name="initialImage" type="impl:NanoXtractEncodedImage"/>
   <element name="npType" type="xsd:string"/>
   <element name="computationalBranches" type="xsd:string"/>
   <element name="nmRatio" type="xsd:double"/>
   <element name="XDSigmaS" type="xsd:double"/>
   <element name="YDSigmaS" type="xsd:double"/>
   <element name="globalThresholdingMethod" type="xsd:string"/>
   <element name="globalThresholdValue" type="xsd:double"/>
   <element name="distanceMergeThreshold" type="xsd:int"/>
   <element name="gaussSize" type="xsd:int"/>
   <element name="sizeMergeThreshold" type="xsd:int"/>
   <element name="minimumSegmentArea" type="xsd:int"/>
   <element name="maximumSegmentArea" type="xsd:int"/>
  </sequence>
 </complexType>
</element>
<complexType name="NanoXtractEncodedImage">
 <sequence>
  <element name="imageFormat" nillable="false" type="xsd:string"/>
  <element name="imageInBase64" nillable="false" type="xsd:string"/>
 </sequence>
</complexType>
```

This has exactly the same input as the Web App. The only difference is that the "image upload" function must be substituted with a complex type of input, the "NanoXtractEncodedImage", that contains two fields. The first field is the format of the image and at this moment NanoXtract supports only two formats, namely "png" and "tif". The other field is the input image (the one that the user uploads in the Web App) encoded in Base64 string. The permissible values for every input element can be seen below:

a) imageFormat can take two values “png” and “tif”

b) npType can be one of “Circular”, Cylindrical”, ‘Plates” and “Other”

c) computationalBranches can be “both” or “bottom”

d) nmRatio > 0

e) 0 ≤ XDSigmasS ≤ 2147483647

f) 0 ≤ YDSigmasS ≤ 2147483647

g) globalThersholdingMethod could be one of: “Huang”, “Intermodes”, “Isodata", “Li”, "Max Entropy”, “Mean”, "Minimum Error”, “Minimum", “Moments", “Otsu", “Percentile", "Renyi Entropy”, “Shanbag", “Triangle", “Yen” and “Manual"

h) 0 ≤ globalThersholdValue ≤ 2147483647

i) 0 ≤ distanceMergeThershold ≤ 2147483647

j) 0 ≤ gaussSize ≤ 2147483647

k) 0 ≤ sizeMergeThershold ≤ 2147483647

l) 0 ≤ minimumSegmentArea ≤ 2147483647

m) 0 ≤ maximumSegmentArea ≤ 2147483647

By calling the NanoXtract Web Service the user receives a response that can be seen in xml format below:

```xml
<complexType name="NanoXtractReply">
 <sequence>
  <element name="successful" type="xsd:boolean"/>
  <element name="errorMessage" nillable="true" type="xsd:string"/>
  <element name="gaussianConvolution" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="globalThresholder" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="waehlbySplitter" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="labelingFilter" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="resultImage" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="resultImageNoOutliers" nillable="true" type="impl:NanoXtractEncodedImage"/>
  <element name="headersMeansDevs" nillable="true" type="impl:HeaderMeanDev"/>
  <element name="schemaValuesPerComponent" nillable="true" type="impl:ArrayOfComponentValues"/>
  <element name="statValuesPerComponent" nillable="true" type="impl:ArrayOfComponentValues"/>
 </sequence>
</complexType>
```

Again, this is exactly the same as the corresponding output of the NanoXtract Web App. The images shown in the Web App are encoded in the Base64 string format. The element “headerMeanDevs” contains the names of the descriptors as well as the mean values of each of the computational branches that have been executed and the corresponding standard deviation. The last two elements contain the corresponding descriptors per component.

The web service can be tested using the free version of SoapUI software. The software can be found and downloaded through the link: https://www.soapui.org/downloads/latest-release.html.

Theoretical descriptor calculation services

Theoretical molecular descriptors are the results of computational procedures, which transform chemical information encoded within a symbolic representation of NMs into numerical features that can complement the experimental characterisation of NMs and can serve as independent variables in the construction of predictive models of NM toxicity for example. A wide range of theoretical descriptors have been proposed for describing the full NP, or addressing only parts of the NM structure such as the core or the coating. The input to the theoretical descriptors calculations tools are molecular file formats such as Simplified Molecular Input Line Entry Specification (SMILES), Structure Data Files (SDF) or crystal structure Protein Data Bank (PDB) formats, to name just a few. The NanoCommons platform will integrate many theoretical descriptor calculation tools. This section of the deliverable report describes and demonstrates the integration of the popular Chemistry Development Kit (CDK) library for cheminformatics and bioinformatics, which has recently been extended to include NM-specific features.

Integration of the Chemistry Development Kit (CDK) library

CDK is a java based descriptor calculation tool which includes mathematical methods and procedures for calculating hundends of topological, geometrical, charge based and constitutional theoretical descriptors using SMILES representations of substances. CDK has been extended to include nano-specific features: The EnergyBandDescriptor and MetalAtomCountDescriptor have been specifically developed for metal oxide NMs. EnergyBandDescriptor uses data from quantum chemical calculations, while the MetalAtomCountDescriptor is based purely on the chemical formula associated with the NM. The ParticleSizeDescriptor is a more general feature, which normalises experimental data into a uniform numerical descriptor. Specifically, the input data may be a range, a mean, etc., while the output must have the same meaning for all materials. The CDK library has been integrated into the Nanocommons platform through the Jaqpot modelling infrastructure.

Integration of the CDK calculation tool in the NanoCommons infrastructure - Availability as a web service

A SMILES descriptor calculation service is also registered in the Jaqpot Platform Descriptor API under the id *'smiles'* and can be accessed programmatically.

Under the operation https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/getDescriptor one can request the service description for the SMILES descriptor calculator with the proper *id* (Figure 44):

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	smiles		path	string

Figure 44: Parameter to GET the smiles service description

The received JSON in response to this GET command is:


```
{
  "meta": {
    "descriptions": [
      "Smiles desc"
    ],
    "titles": [
      "Smiles title"
    ],
    "subjects": [
      "Smiles subject"
    ],
    "creators": [
      "6d9b8261-e255-4fca-9289-0fde48644cf8"
    ],
    "date": "2018-12-18T15:51:01.441+0000",
    "locked": false
  },
  "ontologicalClasses": [
    "ot:Smiles",
    "ot:Algorithm"
  ],
  "parameters": [
    {
      "name": "Categories",
      "value": [
        "all"
      ],
      "scope": "MANDATORY",
      "allowedValues": [
        [
          "all"
        ],
        [
          "topological"
        ],
        [
          "hybrid"
        ],
        [
          "geometrical"
        ],
        [
          "constitutional"
        ],
        [
          "electronic"
        ]
      ],
      "_id": "categories"
    }
  ],
  "ranking": 0,
  "descriptorService": "http://jaqpot-cdk-descriptors.jaqpot:8080/descriptors/smiles/calculate",
  "_id": "smiles"
}
```


It can be seen from the JSON SMILES service description that the supported theoretical descriptor categories are: topological, hybrid, geometrical, constitutional, electronic or all of the above. The set of descriptors extracted for each category are presented in (Steinbeck C. et al., 2006, Willighagen E.L. et al., 2017) and summarized in the following table.

Table 5:

Class	Implemented descriptors
Constitutional	Aromatic atom and bond counts
	Hydrogen bond donor/acceptor counts
	Rotatable bond count
	Proton type
	Pi-contact of two atoms
	Proton RDF
	Rule of Five
	XLogP
Topological	X_t indices ($°X_t$ and ^1X_t)
	X_v indices ($°X_v$ and ^1X_v)
	Wiener number
	Zagreb index
	Vertex adjacency information
	Atomic degree
	Petitjean number
	K shape indices (^1K, ^2K, ^3K)
Geometrical	Gravitational indices

	Shortest path bond count
	Moment of inertia
	Distance in space
Electronic	Sigma electronegativity
	Proton partial charges
	Number of valence electrons
	Polarizability (effective, sum, difference)
Hybrid	BCUT, WHIM
	Topological surface area

The Jaqpot Descriptor API requires the following list of parameters in order to initiate a descriptor calculation procedure:

1. "title": a title for the produced final descriptors Dataset,
2. "description": a description of the dataset that will be returned as the output of the service,
3. "dataset_uri": the specific dataset URI from Jaqpot that will be used to produce the new dataset of descriptors,
4. "description_features": an array of strings specifying the features of the dataset that the service takes as an input (in our case a column that contains base64 images),
5. "parameters": the parameters of the service supplied as an array of values and/or strings.

Below is an example of applying the *smiles* descriptor service to extract *hybrid* descriptors. The csv file containing SMILES representations of the different molecules can be found at the following address: https://github.com/KinkyDesign/descriptor-csv-examples/blob/master/SMILES.csv. The csv file was uploaded in Jaqpot using the procedure described in Annex 1 and produced the following dataset: https://api.jaqpot.org/jaqpot/services/dataset/8899cb23fc2642059c5bd08c8c1292cf. Through the Dataset API, https://api.jaqpot.org/jaqpot/swagger/#!/dataset/getDataset, we can GET it's JSON representation by providing the *id* (Figure 45).

Response Content Type application/json ▼

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	8899cb23fc2642059c5bd08c8c1292cf		path	string
dataEntries	true ▼		query	boolean

Figure 45: Parameter to GET the initial SMILES dataset

The resulting JSON is:


```
{
 "meta": {
  "descriptions": [
   "Smile Dataset without smiles descriptors"
  ],
  "titles": [
   "Initial Smiles Dataset"
  ],
  "creators": [
   "6d9b8261-e255-4fca-9289-0fde48644cf8"
  ],
  "date": "2018-12-29T21:03:24.214+0000",
  "locked": false
 },
 "visible": true,
 "featured": false,
 "dataEntry": [
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row1",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/898b8ce77db641a5a85902439d5a93ab"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL562",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"COC1=CC(=O)C[C@@H](C)[C@]12Oc3c(Cl)c(OC)cc(OC)c3C2=O",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "GRISEOFULVIN"
   },
   "_id": "103c000ce306491595853edac194b876"
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row2",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/4125500ad98144028ba1173410d13a01"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL184",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "Nc1nc(O)c2ncn(COCCO)c2n1",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "ACYCLOVIR"
   },
   "_id": "6adc390417574061ac13faae41533016"
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row3",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
```

"URI": "https://api.jaqpot.org/jaqpot/services/substance/8a2bee65212e4fdfa583029d892a9a1c"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL499",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CC(C)(C)NC[C@H](O)COc1nsnc1N2CCOCC2",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "TIMOLOL"
},
"_id": "5cad628b9c0a4f0c94ac7f9e30f10d26"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row4",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/9c012592054a4d39abe1eff41e5cbcd0"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL546",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CC(C)NCC(O)COc1ccccc1OCC=C",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "OXPRENOLOL"
},
"_id": "77f4111fd08f4b6cbe4faf5faa386607"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row5",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/ff67c275e1934afaa39a1d7c1b5c8b15"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL57",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"Cc1ccnc2N(C3CC3)c4ncccc4C(=O)Nc12",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "NEVIRAPINE"
},
"_id": "f2aa060d75cc4bbfb6a54f9d290ce000"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row6",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/8b5559ca72f94ff29972100de4ff882f"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL3",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "CN1CCC[C@H]1c2cccnc2",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "NICOTINE"
},
"_id": "50e0101db274460f955d9be3f25b45ec"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row7",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/5bc8cdc9818e44d2bc7192547ecd075a"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL1480",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CCOC(=O)C1=C(C)NC(=C(C1c2cccc(Cl)c2Cl)C(=O)OC)C",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "FELODIPINE"
},
"_id": "54a3dc6a42174726a5cfdc5e865237cd"

```
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row8",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/f963e0a1e968447e87b9a9b88d8e0451"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL59",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "NCCc1ccc(O)c(O)c1",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "DOPAMINE"
   },
   "_id": "cc8871515978476dadbda7d16d652cc8"
  }
 ],
 "features": [
  {
   "name": " ChemblID",
   "units": "NA",
   "conditions": {},
   "category": "EXPERIMENTAL",
   "uri": "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad"
  },
  {
   "name": "compound_name",
   "units": "NA",
   "conditions": {},
   "category": "EXPERIMENTAL",
   "uri": "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb"
  },
  {
   "name": " SMILES",
   "units": "NA",
   "conditions": {},
   "category": "EXPERIMENTAL",
   "uri": "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a"
  }
 ],
 "totalRows": 8,
 "totalColumns": 3,
 "_id": "8899cb23fc2642059c5bd08c8c1292cf"
}
```

In the features field we locate the names of the three columns of our dataset, the ChemblId, which contains the Id of the compound in the ChEMBL database, the compound_name, that contains the actual name of the compound, and the SMILES column that contains the SMILES form of the compound. We populate *description_features* with the SMILES feature uri, https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a.

The application of the POST request on the Jaqpot Descriptor service https://api.jaqpot.org/jaqpot/swagger/#!/descriptor/applydescriptor is shown in Figure 46 with the following parameters:

1. “title”: “SMILES dataset”
2. “description”: “SMILES dataset after applying descriptor service”
3. “dataset_uri”: “https://api.jaqpot.org/jaqpot/services/dataset/8899cb23fc2642059c5bd08c8c1292cf”

4. “Description_features”: “https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a”
5. {"categories":["hybrid"]}

Parameters

Parameter	Value	Description	Parameter Type	Data Type
title	SMILES dataset		formData	string
description	SMILES dataset after applying descriptor servic		formData	string
dataset_uri	https://api.jaqpot.org/jaqpot/services/dataset/8899c		formData	string
description_features	https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a		formData	Array[string]
parameters	{"categories":["hybrid"]}		formData	string
id	smiles		path	string
Authorization			header	string

Figure 46: Parameters to POST on the SMILES Descriptor service

The POST operation initiates a descriptor calculation procedure. The user can access information about this task through the Task API, https://api.jaqpot.org/jaqpot/swagger/#!/task/getTask, using the id of the task that is returned from the POST operation. When the asynchronous task is finished successfully, the task reports the id of the resulting dataset. In our case this dataset got the id 05f0d1685fe34d43936b4253ad2df821 and we can GET it from the Dataset API, https://api.jaqpot.org/jaqpot/swagger/#!/dataset/getDataset (Figure 47).

Response Content Type application/json ▼

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	05f0d1685fe34d43936b4253ad2df821		path	string
dataEntries	true ▼		query	boolean

Figure 47: Parameter to GET the resulting SMILES Dataset

The produced JSON representation is shown next:


```
{
 "meta": {
  "comments": [
```

"Created by task TSKdd14072362744e158e7e765e177a6101"
],
"descriptions": [
"SMILES dataset after applying descriptor service"
],
"titles": [
"SMILES dataset"
],
"creators": [
"6d9b8261-e255-4fca-9289-0fde48644cf8"
],
"hasSources": [
"https://api.jaqpot.org/jaqpot/services/dataset/8899cb23fc2642059c5bd08c8c1292cf"
],
"date": "2018-12-29T22:47:06.557+0000",
"locked": false
},
"visible": true,
"dataEntry": [
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "COC1=CC(=O)C[C@@H](C)[C@]12Oc3c(Cl)c(OC)cc(OC)c3C2=O"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "GRISEOFULVIN",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 4.8474698900198545,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d":
"COC1=CC(=O)C[C@@H](C)[C@]12Oc3c(Cl)c(OC)cc(OC)c3C2=O",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL562",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.3666691011668108,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.988451375604495,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 34.96885373839483,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 12.203366922467115,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.28403007950072084
},
"_id": "51a5105118a942a2b65d2bc9bbbde313"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "Nc1nc(O)c2ncn(COCCO)c2n1"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "ACYCLOVIR",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 3.811793457165341,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d": "Nc1nc(O)c2ncn(COCCO)c2n1",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL184",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.3934629549099172,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.988488120784117,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 15.996928323467563,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 9.173245473922986,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.25155538340366673
},
"_id": "acf5aa6a01e847a7add5e3f2d6101c23"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "CC(C)(C)NC[C@H](O)COc1nsnc1N2CCOCC2"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "TIMOLOL",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 5.230817892142408,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d":
"CC(C)(C)NC[C@H](O)COc1nsnc1N2CCOCC2",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL499",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.38896519576529565,

"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.99436975153516,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 31.972072062909024,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 10.226007459851793,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.21506401178675807
},
"_id": "2f5f0c36cf804081b986b33102696f95"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "CC(C)NCC(O)COc1ccccc1OCC=C"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "OXPRENOLOL",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 5.061597460385779,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d": "CC(C)NCC(O)COc1ccccc1OCC=C",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL546",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.3889006868095421,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.790000000000001,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 15.996927517698522,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 8.826241701978692,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.1282047129074541
},
"_id": "ecd8ef4e55e44424be420cba1d39a762"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "Cc1cccnc2N(C3CC3)c4ncccc4C(=O)Nc12"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "NEVIRAPINE",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 5.390164643918201,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d":
"Cc1ccnc2N(C3CC3)c4ncccc4C(=O)Nc12",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL57",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.27453230695317216,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.850000000000001,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 15.994920434000875,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 11.238949747349041,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.252229350860743
},
"_id": "bda35d74ce4e43c695a6e5846627ac57"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "CN1CCC[C@H]1c2cccnc2"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "NICOTINE",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 6.259676733770103,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d": "CN1CCC[C@H]1c2cccnc2",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL3",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.298916454147278,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.850000000000001,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 14.00408418640803,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 9.725070244123906,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.16149147523303153
},
"_id": "5f05d7fd2f174043befe41ddd7401c08"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "CCOC(=O)C1=C(C)NC(=C(C1c2cccc(Cl)c2Cl)C(=O)OC)C"
},
"values": {

"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "FELODIPINE",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 5.054456292718001,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d":
"CCOC(=O)C1=C(C)NC(=C(C1c2cccc(Cl)c2Cl)C(=O)OC)C",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL1480",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.31566476561677576,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.890000000000008,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 34.96985476718037,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 13.450537982149193,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.29236686373851584
},
"_id": "4a1824f4ffbc4c36b15d6e8d59cfe59c"
},
{
"datasetId": "05f0d1685fe34d43936b4253ad2df821",
"entryId": {
"name": "NCCc1ccc(O)c(O)c1"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1": "DOPAMINE",
"https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9": 4.846714244375004,
"https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d": "NCCc1ccc(O)c(O)c1",
"https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b": "CHEMBL59",
"https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200": -0.3587508979659609,
"https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a": 11.850000000000003,
"https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec": 15.995919677679218,
"https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f": 8.602836377027094,
"https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932": 0.11570186518560004
},
"_id": "95b74de058a5484c967691f43cc21352"
}
],
"features": [
{
"name": "compound_name",
"units": "NA",
"conditions": {},
"category": "EXPERIMENTAL",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/01b85fceb7f945498df83703b51ecad1"
},
{
"name": "BCUTp-1h",
"category": "CDK",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/94880f6d8d9843e3a47017c404b0697f"
},
{
"name": "BCUTc-1l",
"category": "CDK",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/5106c755a68145698e1afdb15feae200"
},
{
"name": "BCUTc-1h",
"category": "CDK",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/d947d26a6b9046bd92e992cc94b23932"
},
{
"name": " ChemblID",
"units": "NA",
"conditions": {},
"category": "EXPERIMENTAL",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/21b16299cc0c490c93c406f5e9e52e9b"
},
{
"name": "BCUTw-1h",
"category": "CDK",
"uri": "https://api.jaqpot.org/jaqpot/services/feature/91a4ccd515e34b48bb57b5cb887cccec"
},
{

```
    "name": "BCUTw-1l",
    "category": "CDK",
    "uri": "https://api.jaqpot.org/jaqpot/services/feature/7789c1197a3247699a9baa28ee2a4e6a"
  },
  {
    "name": " SMILES",
    "units": "NA",
    "conditions": {},
    "category": "EXPERIMENTAL",
    "uri": "https://api.jaqpot.org/jaqpot/services/feature/212d757662fd477c98b808903d28225d"
  },
  {
    "name": "BCUTp-1l",
    "category": "CDK",
    "uri": "https://api.jaqpot.org/jaqpot/services/feature/1d1b52c2b01740fd95b4a49401b6b6a9"
  }
 ],
 "totalRows": 8,
 "totalColumns": 9,
 "descriptors": [
  "CDK"
 ],
 "existence": "DESCRIPTORSADDED",
 "_id": "05f0d1685fe34d43936b4253ad2df821"
}
```

Evaluation of physicochemical properties of NMs

NMs are often characterized by their physicochemical properties where experimental data are commonly used to create a relationship between the biological activity of NMs and their physicochemical properties. Furthermore, the toxicity of NMs is found to be related to their molecular and electronic properties, especially those determining the nanobio interactions. Quantitative Structure Activity Relationships (QSAR), read-across and grouping methods are used to establish such relationships. The bionano interactions are not easily accessible experimentally due to complexity and diversity of the systems in question and most important relevant properties responsible for NM transport and biological activity are not routinely reported. Therefore, for construction of predictive models it is essential to get access to physicochemical properties of NMs through theoretical calculations (materials modelling) to complement experimental or theoretical data from the literature. Our ultimate goal is to collect and integrate into the NanoCommons KB modelling and computational tools which are most efficient, i.e. computationally feasible (fast and accurate methods with a low computational cost), capable to research a broad range of NMs, e.g., regarding their size, chemical type and physical form as well as applicability. The outcome of the simulations of the modelling tools will serve to establish a structure activity (toxicity) relationship. Finally, all data will be gathered and systematically categorized and annotated to be integrated into the KB.

Molecular and electronic properties

The software for the effective theoretical calculation of the molecular and electronic properties of NMs ought to cover a broad size range of NMs and therefore semi-empirical calculations with the MOPAC program package will be an essential tool to describe the microscopic quantum mechanical nature of NMs. The molecular and electronic properties obtained from semi-empirical calculations can be cross-validated by higher level of accuracy calculations, mainly, through density functional theory (DFT) methods and the NWChem program package. For a few properties, e.g. hydration free energy, molecular adsorption, van der Waals interactions, molecular dynamics (MD) simulations with e.g. GROMACS are prevalent. Furthermore, we will address many bionano interaction properties, such as protein binding affinity ranking with statistical models.

A brief description of the software used to evaluate the NM molecular descriptors is given below.

1. MOPAC:

MOPAC is a computer program which is widely used in the calculation of the geometries and energies of biological systems (proteins), crystals, co-crystals, polymers and other condensed phase systems. MOPAC 2016 has been improved with regards to the description of intermolecular interaction. Also, fast property prediction algorithms for screening libraries of drug-sized molecules for a wide range of properties including pKa are implemented (Stewart, 2007; Stewart, 2016). Thus, the generation of electronic descriptors for QSARs is the canonical route. The effect of solvents can be captured by the COSMO method (Klamt et al., 1993) where both ground and excited state geometries of solvated systems are optimized.

2. NWChem:

NWChem is a modern computational chemistry program package with state-of-the-art methodology from quantum to classical and all combinations implemented for the efficient calculation of a variety of molecular and electronic properties of biomolecules, nanostructures and solid-state systems, utilizing parallel computing resources most efficiently (Valiev et al., 2010). Noteworthy are the multiple capabilities in the areas of kinetics and dynamics of chemical transformations, chemistry at interfaces and in the condensed phase.

3. SIESTA:

SIESTA is a program package and a method which is used to perform efficient electronic structure calculations and *ab initio* MD simulations of molecules and solids (Soler et al., 2002). Localized basis sets and linear-scaling algorithms feature fast and accurate simulations. This approach projects the electron wavefunctions and density onto a real-space grid. It is written in Fortran 95 and memory is allocated dynamically. It may be compiled for serial or parallel execution (under Message Passing Interface, MPI). In particular, layers of NPs of a specific size using periodic boundary conditions can be modelled and treated with density functional methods.

4. GROMACS:

GROMACS is a versatile package to perform MD simulation and energy minimization (Hess et al., 2008; Abraham et al., 2015). The Newtonian equations of motion for systems with hundreds, to millions of particles are solved. GROMACS introduces hybrid acceleration by making use of GPUs to accelerate non-bonded force calculation.

5. DFTB+:

DFTB+ is a quantum mechanical simulation software package (Aradi et al., 2007). It is designed to run quantum mechanical simulations in an approximate way to DFT. The dynamics and electronic structure can be studied in molecules, solutions and solid-state materials. DFTB+ is free software licensed under the *GNU Lesser General Public License*.

Other software:

Additional NM properties can be evaluated using free simulation packages OpenMM, LAMMPS, GPAW. In Table 6 all the descriptors are listed along with the methodology and software to be applied as well as the class of NMs that have/can be addressed and the nm size range which is reported in previous publications. The majority of the descriptors (band gap, ionization potential, density of states, electronegativity, etc.) in Table 6 will be evaluated by quantum mechanical methods and are therefore classified as intrinsic properties.

Interface properties

1. ESPResSo:

ESPResSo is a highly versatile software package for performing and analyzing MD many-particle simulations of coarse-grained atomistic or bead-spring models as they are used in soft matter research in physics, chemistry and molecular biology (Limbach et al., 2006; Arnold et al., 2012). It can be used to simulate systems such as polymers, liquid crystals, colloids, polyelectrolytes, and biological systems, such as proteins, DNA, and lipid membranes. It also has a dissipative particle dynamics and lattice Boltzmann solver for hydrodynamic interactions, and allows several particle couplings to the Lipid Bilayer fluid. ESPResSo is free software licensed under the *GNU Lesser General Public License*.

2. SmartNanoTox modelling tools:

SmartNanoTox modelling tools (SNT-MT) are developed by H2020 project SmartNanoTox for prediction of the biological activity of NMs (Lopez et. al, 2015; Brandt et al. 2015; Lopez et al. 2017). The tools include NM and biomolecule coarse-graining using Python scripts, evaluation of parameters of interactions such as potentials of mean force for biomolecular segments at the surface of NMs (using freeware Gromacs package), and prediction of protein 3D structure using free iTasser tool (Zhang et al., 2008). The considered biomolecular segments are as follows:

- Amino acids side chain analogues. There are 20 naturally occurring amino acids from which all proteins of living organisms consist. Each full amino acid contains a peptide backbone fragment which is common for all amino acids. In order to avoid redundancy, we excluded the backbone fragment and considered the side chain analogues for all amino acids excluding glycine (for which side chain analogue is just a H atom) and proline which has a different structure. This set of side chain analogues consists of 18 molecules.
- Glycine and proline amino acids with the backbone fragment (GLY, PRO).
- Modified charged forms of amino acids having pKa values between 4 and 10 (HIS+, GAN, CYS-)
- Molecules representing fragments of the most abundant lipid types: choline (CHL) and phosphate (PHO) group of phosphatydilcholine (PC) lipids; amino group (ETA) of etanolamine lipids (PE), ester group (EST).
- d-glucose as representing sugars.

The above list of molecules covers all the main types of chemical entities: hydrophobic, polar, aromatic and charged, and represents well all typical molecular fragments present in biofluids. A set of binding free energies of these molecules constitutes a simplest "fingerprint" of how an organism sees the NM and thus determines the biological response to a NM.

Other descriptors include overall protein or lipid adsorption free energies for a range of NM sizes 1-100 nm (spheres and cylinders), and binding affinity rankings. The adsorption free energies are evaluated at the coarse-grained level using united atom model, using a SmartNanoTox multiscale technique as outlined in Fig. 48. An example of protein ranking by the binding affinity to a rutile TiO_2 NP is shown in Fig. 49.

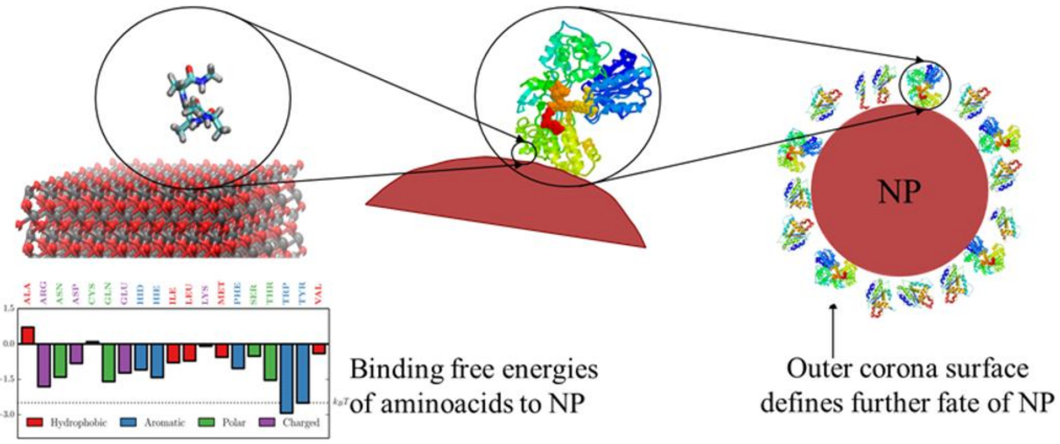


Figure 48: Schematic of the SmartNanoTox scheme for evaluation of biomolecule adsorption free energies on NPs.

Increasing affinity ←

NP Radius, nm	Rank					
	1	2	3	4	5	6
5	A2M	IgG	Fib	Tra	A1A	HSA
20	Fib	A2M	IgG	Tra	A1a	HSA
50	Fib	A2M	IgG	Tra	A1A	HSA
100	Fib	IgG	A2M	Tra	A1A	HSA
500	Fib	IgG	A2M	Tra	A1A	HSA
	Fib	A2M	IgG	A1A	Tra	HSA

Large Small

Figure 49: Example of protein ranking by the binding affinity to a rutile TiO_2 NP.

SNT-MT are open source free software licensed under the *GNU Lesser General Public License* (http://www.smartnanotox.eu/?page_id=143).

Table 6: Methodology and software tools for the calculation of theoretical descriptors of NMs

Descriptor name	Class of materials	Method	Simulation package	NM size range in nm	References
Band gap	MO, CNT	PM6, DFT	MOPAC, NWChem	20.0-30.0 15.0-90.0 0.5-5.0 10.0-70.0 15.0-90.0 15.0-90.0 <5.0 15.0-90.0 15.0-90.0	Burello et al., 2011 Puzyn et al., 2011 Gajewicz et al., 2011 Liu et al., 2013 Mu et al., 2016 Gajewicz et al., 2015 Bae et al., 2015 Mikolajczyk et al., 2015 Pathakoti et al., 2014
Ionization potential	TM, MO, CNT	PM6, HF, DFT, TDDFT	MOPAC, NWChem	15.0-90.0	Puzyn et al., 2011
Density of states	TM, MO, CNT	PM6, HF, DFT	MOPAC, NWChem		
Hydrophobicity	MO, HCNP	CG	SNT-MT	5.0-100.0	Lopez et al., 2015, 2017 Brandt et al., 2015
Dipole moment	MO, CNT	PM6 (COSMO), DFT	MOPAC, NWChem	0.5-5.0	Gajewicz et al., 2011
Atomic charge	TM, MO, CNT	PM6, DFT	MOPAC, NWChem		
Total charge	TM, MO, CNT	PM6, DFT	MOPAC, NWChem		
Electronegativity	MO	PM6, DFT, LDM	MOPAC, NWChem	15.0-150.0 15.0-90.0	Sizochenko et al., 2014 Mu et al., 2016 Sizochenko et al., 2015
Surface charge	TM, MO			15.0-90.0 15.0-90.0 14.0-144.0	Gajewicz et al., 2015 Mikolajczyk et al., 2015 Wyrzykowska et al., 2016
Absorption spectra	TM, MO	CIS, TDDFT	MOPAC, NWChem	<5.0	Bae et al., 2015
Binding energies	TM, MO, CNT	PM6, DFT, DFTB	MOPAC, GPAW, DFTB+		Redel et al., 2009
vdW energy	Spherical NP	CG, PM7, DFT	SNT-MT, MOPAC,	5.0-100.0	Lopez et al., 2015, 2017 Brandt et al., 2015

			NWChem, SIESTA		Luo et al., 2014
Hydration free energies		Integral equation theory			Ratkova et al., 2010
Corona kinetics	TM, MO	CG	ESPResSo	100.0	Lopez et al., 2015, 2017 Brandt et al., 2015
Binding affinities	HCNP	CG	SNT-MT	5.0-100.0	Lopez et al., 2015, 2017 Brandt et al., 2015
Hamaker constants	MO, TM	CG, Lifschitz theory	SNT-MT	5.0-100.0	Lopez et al., 2015, 2017 Brandt et al., 2015 Faure et al., 2011 Pinchuk et al., 2012
Adsorption energies for proteins	HCNP	CG, DFT, DFTB	SNT-MT, GROMACS, SIESTA, DFTB+	5.0-100.0	Lopez et al., 2015, 2017 Brandt et al., 2015
Vibrational frequencies		PM6, MD	MOPAC, GROMACS		

Conclusions

In this deliverable we have described and demonstrated the tools that have been developed and have been included into the NanoCommons modelling infrastructure during the first 12 months of the project for processing raw data and for calculating theoretical descriptors. Image analysis tools have been demonstrated to automatically compute NM descriptors from raw data contained in electronic images. The tools address the specific requirements of various shapes of NMs including spherical, tube-shaped/cylindrical and plates, as well as the complex morphologies resulting from environmental ageing of NMs. The CDK open-source software has also been integrated as a web service for the calculation of various descriptors that characterise the core, the coatings or the full NM. The tools have been developed as interoperable, standards-compliant modular web services, able to maximise cross-talk and interaction between different/diverse sources of data. The set of bionano interaction tools developed in SmartNanoTox are also presented in this deliverable and are available for NanoCommons TA. The tools will gradually be integrated directly with the NanoCommons Knowledgebase such that all data generated will be captured and stored in the database. A list of software tools that can be used for computing molecular, electronic and interfacial properties of NMs was provided at the end of the document. The library of tools will constantly be extended and updated during the course of the project to address the needs of the nanosafety community and will include additional services provided by project partners and third parties as they become available.

References

Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *The Journal of Physical Chemistry A* **2007**, *111*, 5678–5684.

Arnold, A.; Lenz, O.; Kesselheim, S.; Weeber, R.; Fahrenberger, F.; Roehm, D.; Košovan, P.; Holm, C. ESPResSo 3.1: Molecular Dynamics Software for Coarse-Grained Models. *Meshfree Methods for Partial Differential Equations VI Lecture Notes in Computational Science and Engineering* **2012**, 1–23.

Bae, G.-T.; Aikens, C. M. Time-Dependent Density Functional Theory Studies of Optical Properties of Au Nanoparticles: Octahedra, Truncated Octahedra, and Icosahedra. *The Journal of Physical Chemistry C* **2015**, *119*, 23127–23137.

Brandt, E. G.; Lyubartsev, A. P. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO2 (100) Surface. *The Journal of Physical Chemistry C* **2015**, *119*, 18126–18139.

Burello, E.; Worth, A. P. A Theoretical Framework for Predicting the Oxidative Stress Potential of Oxide Nanoparticles. *Nanotoxicology* **2011**, *5*, 228–235.

Doganis P.; Tsiliki, G.; Sarimveis, H.; Chomenidis, H.; Drakakis, G.; Willighagen, E.; Hardy, B. eNanoMapper Deliverable Report D4.2: Descriptor Calculation Algorithms and Methods. *Zenodo* **2015,** http://doi.org/10.5281/zenodo.375609.

Olson E. Particle shape factors and their use in image analysis - Part-1: theory. J GXP Compliance **2011**, 15(3), 85–96.

Faure, B.; Salazar-Alvarez, G.; Bergström Lennart. Hamaker Constants of Iron Oxide Nanoparticles. *Langmuir* **2011**, *27*, 8659–8664.

Ferreira, W. R. T. *IJ 1.46r*, **2012**, 185.

Gajewicz, A.; Puzyn, T.; Rasulev, B.; Leszczynska, D.; Leszczynski, J. Metal Oxide Nanoparticles: Size-Dependence of Quantum-Mechanical Properties. *Nanoscience & Nanotechnology-Asia* **2011**, *1*, 53–58.

Gajewicz, A.; Schaeublin, N.; Rasulev, B.; Hussain, S.; Leszczynska, D.; Puzyn, T.; Leszczynski, J. Towards Understanding Mechanisms Governing Cytotoxicity of Metal Oxides Nanoparticles: Hints from Nano-QSAR Studies. *Nanotoxicology* **2015**, *9*, 313–325.

GitHub ImageJ1, Ellipse Fitter, https://github.com/imagej/imagej1/blob/master/ij/process/EllipseFitter.java, (accessed 9 October 2018).

Glasbey, C. An Analysis of Histogram-Based Thresholding Algorithms. *CVGIP: Graphical Models and Image Processing* **1993**, *55*, 532–537.

Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

Huang, L.-K.; Wang, M.-J. J. Image Thresholding by Minimizing the Measures of Fuzziness. *Pattern Recognition* **1995**, *28*, 41–51.

ImageJ, https://github.com/imagej/imagej-ops/find/master, (accessed 19 January 2018).

ImageJ Shape Filter, https://imagej.net/Shape_Filter, (accessed 9 October 2018).

Kapur, J.; Sahoo, P.; Wong, A. A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Computer Vision, Graphics, and Image Processing* **1985**, *29*, 273–285.

Klamt, A.; Schüürmann, G. COSMO: a New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

Komsta, L. *R Packag. version 0.14*, **2011**, . http://CRAN.R-project.org/package=outliers.

Limbach, H.; Arnold, A.; Mann, B.; Holm, C. ESPResSo—an Extensible Simulation Package for Research on Soft Matter Systems. *Computer Physics Communications* **2006**, *174*, 704–727.

Liu, R.; Zhang, H. Y.; Ji, Z. X.; Rallo, R.; Xia, T.; Chang, C. H.; Nel, A.; Cohen, Y. Development of Structure–Activity Relationship for Metal Oxide Nanoparticles. *Nanoscale* **2013**, *5*, 5644-5653.

Lopez, H.; Brandt, E. G.; Mirzoev, A.; Zhurkin, D.; Lyubartsev, A.; Lobaskin, V. Multiscale Modelling of Bionano Interface. *Advances in Experimental Medicine and Biology Modelling the Toxicity of Nanoparticles* **2017**, 173–206.

Lopez, H.; Lobaskin, V. Coarse-Grained Model of Adsorption of Blood Plasma Proteins onto Nanoparticles. *The Journal of Chemical Physics* **2015**, *143*, 243138.

Luo, Y.; Zhao, R.; Pendry, J. B. Van Der Waals Interactions at the Nanoscale: The Effects of Nonlocality. *Proceedings of the National Academy of Sciences* **2014**, *111*, 18422–18427.

Melagraki, G.; Afantitis, A. Enalos InSilicoNano Platform: an Online Decision Support Tool for the Design and Virtual Screening of Nanoparticles. *RSC Adv.* **2014**, *4*, 50713–50725.

Mikolajczyk, A.; Gajewicz, A.; Rasulev, B.; Schaeublin, N.; Maurer-Gardner, E.; Hussain, S.; Leszczynski, J.; Puzyn, T. Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure–Property Relationship Approach. *Chemistry of Materials* **2015**, *27*, 2400–2407.

Mikolajczyk, A.; Gajewicz, A.; Rasulev, B.; Schaeublin, N.; Maurer-Gardner, E.; Hussain, S.; Leszczynski, J.; Puzyn, T. Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure–Property Relationship Approach. *Chemistry of Materials* **2015**, *27*, 2400–2407.

Mondini, S.; Ferretti, A. M.; Puglisi, A.; Ponti, A. Pebbles and PebbleJuggler: Software for Accurate, Unbiased, and Fast Measurement and Analysis of Nanoparticle Morphology from Transmission Electron Microscopy (TEM) Micrographs. *Nanoscale* **2012**, *4*, 5356.

Mu, Y.; Wu, F.; Zhao, Q.; Ji, R.; Qie, Y.; Zhou, Y.; Hu, Y.; Pang, C.; Hristozov, D.; Giesy, J. P.; *et al.*Predicting Toxic Potencies of Metal Oxide Nanoparticles by Means of Nano-QSARs. *Nanotoxicology* **2016**, *10*, 1207–1214.

Odziomek, K.; Ushizima, D.; Oberbek, P.; Kurzydłowski, K. J.; Puzyn, T.; Haranczyk, M. Scanning Electron Microscopy Image Representativeness: Morphological Data on Nanoparticles. *Journal of Microscopy* **2017**, *265*, 34–50.

Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9*, 62–66.

Pathakoti, K.; Huang, M.-J.; Watts, J. D.; He, X.; Hwang, H.-M. Using Experimental Data of Escherichia Coli to Develop a QSAR Model for Predicting the Photo-Induced Cytotoxicity of Metal Oxide Nanoparticles. *Journal of Photochemistry and Photobiology B: Biology* **2014**, *130*, 234–240.

Pinchuk, A. O. Size-Dependent Hamaker Constant for Silver Nanoparticles. *The Journal of Physical Chemistry C* **2012**, *116*, 20099–20102.

Prewitt, J.M.S.; Mendelsohn, M.L. The analysis of cell images. *Annals of the New York Academy of Sciences* **1966**, 128, 1035-1053.

Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using Nano-QSAR to Predict the Cytotoxicity of Metal Oxide Nanoparticles. *Nature Nanotechnology* **2011**, *6*, 175–178.

Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. *The Journal of Physical Chemistry B* **2010**, *114*, 12068–12079.

Redel, E.; Walter, M.; Thomann, R.; Vollmer, C.; Hussein, L.; Scherer, H.; Krüger, M.; Janiak, C. Synthesis, Stabilization, Functionalization and, DFT Calculations of Gold Nanoparticles in Fluorous Phases (PTFE and Ionic Liquids). *Chemistry - A European Journal* **2009**, *15*, 10047–10059.

Richard, F. **2007**, Moment Calculator ImageJ plugin documentation. https://imagej.nih.gov/ij/plugins/moments.html

Ridler, T.W.; Calvard, S. Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man, and Cybernetics* **1978**, *8*, 630–632.

Rueden, C. T.; Schindelin, J.; Hiner, M. C.; Dezonia, B. E.; Walter, A. E.; Arena, E. T.; Eliceiri, K. W. ImageJ2: ImageJ for the next Generation of Scientific Image Data. *BMC Bioinformatics* **2017**, *18(1):529*.

Schindelin, J.; Hiner, M.; Helfrich, S. **2017**, ImageJ AutoThreshold https://imagej.net/Auto_Threshold

Sezgin, M.; Sankur, B. Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. *Journal of Electronic Imaging* **2004**, *13*, 146-165.

Shanbhag, A. Utilization of Information Measure as a Means of Image Thresholding. *CVGIP: Graphical Models and Image Processing* **1994**, *56*, 414–419.

Sizochenko, N.; Rasulev, B.; Gajewicz, A.; Kuz'min, V.; Puzyn, T.; Leszczynski, J. From Basic Physics to Mechanisms of Toxicity: the "Liquid Drop" Approach Applied to Develop Predictive Classification Models for Toxicity of Metal Oxide Nanoparticles. *Nanoscale* **2014**, *6*, 13986–13993.

Sizochenko, N.; Rasulev, B.; Gajewicz, A.; Mokshyna, E.; Kuz'min, V. E.; Leszczynski, J.; Puzyn, T. Causal Inference Methods to Assist in Mechanistic Interpretation of Classification Nano-SAR Models. *RSC Advances* **2015**, *5*, 77739–77745.

Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA Method For ab Initio order-N materials Simulation. *Journal of Physics: Condensed Matter* **2002**, *14*, 2745–2779.

Steger, C. An Unbiased Detector of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1998**, *20*, 113–125.

Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. Recent Developments of the Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics. *Current Pharmaceutical Design* **2006**, *12*, 2111–2120.

Stewart, J. J. P. MOPAC2016: Stewart Computational Chemistry, Colorado Springs, CO, USA, **2016**, http://OpenMOPAC.net.

Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *Journal of Molecular Modeling* **2007**, *13*, 1173–1213.

Toth, P.; Farrer, J.; Palotas, A.; Lighty, J.; Eddings, E. Automated Analysis of Heterogeneous Carbon Nanostructures by High-Resolution Electron Microscopy and on-Line Image Processing. *Ultramicroscopy* **2013**, *129*, 53–62.

Tsai, W.-H. Moment-Preserving Thresolding: A New Approach. *Computer Vision, Graphics, and Image Processing* **1985**, *29*, 377–393.

Valiev, M.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; Dam, H. V.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T.; *et al.*NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. *Computer Physics Communications* **2010**, *181*, 1477–1489.

Varsou, D.-D.; Afantitis, A.; Tsoumanis, A.; Melagraki, G.; Sarimveis, H.; Valsami-Jones, E.; Lynch, I. A Safe-by-Design Tool for Functionalised Nanomaterials through the Enalos Nanoinformatics Cloud Platform. *Nanoscale Advances* **2019**.

Vilanova, O.; Mittag, J. J.; Kelly, P. M.; Milani, S.; Dawson, K. A.; Rädler, J. O.; Franzese, G. Understanding the Kinetics of Protein–Nanoparticle Corona Formation. *ACS Nano* **2016**, *10*, 10842–10850.

Walkey, C. D.; Olsen, J. B.; Song, F.; Liu, R.; Guo, H.; Olsen, D. W. H.; Cohen, Y.; Emili, A.; Chan, W. C. W. Protein Corona Fingerprinting Predicts the Cellular Interaction of Gold and Silver Nanoparticles. *ACS Nano* **2014**, *8*, 2439–2455.

Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Chertó, M.; Spjuth, O.; Torrance, G.; Evelo, C.T.; Guha, R.; Steinbeck,C. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching, *Journal of Cheminformatics* **2017,** 9:33

Wyrzykowska, E.; Mikolajczyk, A.; Sikorska, C.; Puzyn, T. Development of a Novel in Silico Model of Zeta Potential for Metal Oxide Nanoparticles: a Nano-QSPR Approach. *Nanotechnology* **2016**, *27*, 445702.

Mingqiang, Y.; Kidiyo, K.; Joseph, R. A Survey of Shape Feature Extraction Techniques. Pattern Recognition Techniques, Technology and Applications **2008**.

Zack, G. W.; Rogers, W. E.; Latt, S. A. Automatic Measurement of Sister Chromatid Exchange Frequency. *Journal of Histochemistry & Cytochemistry* **1977**, *25*, 741–753.

Zhang, P.; Ma, Y.; Zhang, Z.; He, X.; Zhang, J.; Guo, Z.; Tai, R.; Zhao, Y.; Chai, Z. Biotransformation of Ceria Nanoparticles in Cucumber Plants. *ACS Nano* **2012**, 6, 9943–9950.

Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9*, 40.

Appendix 1

Importing csv files in the Jaqpot platform

Jaqpot supports different ways of importing data into the system. One straightforward way is via the upload of .csv files using Jaqpot's Dataset API.

More specifically, under the https://api.jaqpot.org/jaqpot/swagger/#/dataset API there is the *POST /dataset/csv* operation. The different parameters required by the method are the following:

1. "title": the title of the resulting dataset
2. "description": a description of the content of the resulting dataset
3. "file": the .csv file to upload

The *file* field accepts all kinds of .csv files, however, in order for the operation to complete successfully some general rules should be followed when creating the .csv file, as follows:

- The first row of the .csv file must contain the name of each column. This populates the *feature* names of the resulting dataset;
- The service understands the default csv value separator, comma. A full stop or point is accepted as a decimal separator in numerical values.
- Numerical values should not be enclosed in quotes.

Below is an example of importing a csv file into the system:

We will use https://github.com/KinkyDesign/descriptor-csv-examples/blob/master/SMILES.csv as the example csv to import.

```
compound_name, ChemblID, SMILES
"GRISEOFULVIN","CHEMBL562","COC1=CC(=O)C[C@@H](C)[C@]12Oc3c(Cl)c(OC)cc(OC)c3C2=O"
"ACYCLOVIR","CHEMBL184","Nc1nc(O)c2ncn(COCCO)c2n1"
"TIMOLOL","CHEMBL499","CC(C)(C)NC[C@H](O)COc1nsnc1N2CCOCC2"
"OXPRENOLOL","CHEMBL546","CC(C)NCC(O)COc1ccccc1OCC=C"
"NEVIRAPINE","CHEMBL57","Cc1ccnc2N(C3CC3)c4ncccc4C(=O)Nc12"
"NICOTINE","CHEMBL3","CN1CCC[C@H]1c2cccnc2"
"FELODIPINE","CHEMBL1480","CCOC(=O)C1=C(C)NC(=C(C1c2cccc(Cl)c2Cl)C(=O)OC)C"
"DOPAMINE","CHEMBL59","NCCc1ccc(O)c(O)c1"
```

As indicated, all entries, except from the first row, are enclosed in quotes. That is because we do not have any numerical values in this example.

We fill the POST parameters with the following values as shown in Figure A1:

1. "title": "Sample SMILES Dataset"

2. “description”: “Dataset of 8 compounds with SMILES specifications”
3. “file”: SMILES.csv

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
file	Choose File SMILES.csv	xls[m,x] file	formData	file
title	Sample SMILES Dataset	Title of dataset	formData	string
description	Dataset of 8 compounds with SMILES specifica	Description of dataset	formData	string

Try it out!

Figure A1: Parameters to POST on dataset/csv service

If the service is successful it returns the dataset description (dataset without the dataEntries). We can receive the full dataset by GETting the dataset through the https://api.jaqpot.org/jaqpot/swagger/#!/dataset/getDataset method, using the id of the dataset, which is provided in the dataset description. In our example, the dataset id is *8899cb23fc2642059c5bd08c8c1292cf* (Figure A2).

Parameters

Parameter	Value	Description	Parameter Type	Data Type
Authorization		Authorization token	header	string
id	8899cb23fc2642059c5bd08c8c1292cf		path	string
dataEntries	true ▼		query	boolean

Figure A2: Parameters to GET the imported Dataset

The produced Jaqpot dataset in JSON format is:


```
{
 "meta": {
  "descriptions": [
   "Smile Dataset without smiles descriptors"
  ],
  "titles": [
   "Initial Smiles Dataset"
  ],
  "creators": [
   "6d9b8261-e255-4fca-9289-0fde48644cf8"
  ],
  "date": "2018-12-29T21:03:24.214+0000",
  "locked": false
 },
 "visible": true,
 "featured": false,
```

```
 "dataEntry": [
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row1",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/898b8ce77db641a5a85902439d5a93ab"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL562",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"COC1=CC(=O)C[C@@H](C)[C@]12Oc3c(Cl)c(OC)cc(OC)c3C2=O",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "GRISEOFULVIN"
   },
   "_id": "103c000ce306491595853edac194b876"
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row2",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/4125500ad98144028ba1173410d13a01"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL184",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "Nc1nc(O)c2ncn(COCCO)c2n1",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "ACYCLOVIR"
   },
   "_id": "6adc390417574061ac13faae41533016"
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row3",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/8a2bee65212e4fdfa583029d892a9a1c"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL499",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CC(C)(C)NC[C@H](O)COc1nsnc1N2CCOCC2",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "TIMOLOL"
   },
   "_id": "5cad628b9c0a4f0c94ac7f9e30f10d26"
  },
  {
   "datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
   "entryId": {
    "name": "row4",
    "ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
    "URI": "https://api.jaqpot.org/jaqpot/services/substance/9c012592054a4d39abe1eff41e5cbcd0"
   },
   "values": {
    "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL546",
    "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CC(C)NCC(O)COc1ccccc1OCC=C",
    "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "OXPRENOLOL"
   },
   "_id": "77f4111fd08f4b6cbe4faf5faa386607"
```

},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row5",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/ff67c275e1934afaa39a1d7c1b5c8b15"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL57",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"Cc1ccnc2N(C3CC3)c4ncccc4C(=O)Nc12",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "NEVIRAPINE"
},
"_id": "f2aa060d75cc4bbfb6a54f9d290ce000"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row6",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/8b5559ca72f94ff29972100de4ff882f"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL3",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "CN1CCC[C@H]1c2cccnc2",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "NICOTINE"
},
"_id": "50e0101db274460f955d9be3f25b45ec"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row7",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/5bc8cdc9818e44d2bc7192547ecd075a"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL1480",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a":
"CCOC(=O)C1=C(C)NC(=C(C1c2cccc(Cl)c2Cl)C(=O)OC)C",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "FELODIPINE"
},
"_id": "54a3dc6a42174726a5cfdc5e865237cd"
},
{
"datasetId": "8899cb23fc2642059c5bd08c8c1292cf",
"entryId": {
"name": "row8",
"ownerUUID": "7da545dd-2544-43b0-b834-9ec02553f7f2",
"URI": "https://api.jaqpot.org/jaqpot/services/substance/f963e0a1e968447e87b9a9b88d8e0451"
},
"values": {
"https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad": "CHEMBL59",
"https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a": "NCCc1ccc(O)c(O)c1",
"https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb": "DOPAMINE"
},
"_id": "cc8871515978476dadbda7d16d652cc8"
}

```
  ],
  "features": [
    {
      "name": " ChemblID",
      "units": "NA",
      "conditions": {},
      "category": "EXPERIMENTAL",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/_ChemblID_00867b85decb42f89b935f3434c423ad"
    },
    {
      "name": "compound_name",
      "units": "NA",
      "conditions": {},
      "category": "EXPERIMENTAL",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/compound_name_e44910d667ef4ee0805d850a096826cb"
    },
    {
      "name": " SMILES",
      "units": "NA",
      "conditions": {},
      "category": "EXPERIMENTAL",
      "uri": "https://api.jaqpot.org/jaqpot/services/feature/_SMILES_650a25bc84df4adfa7cea0503cb4282a"
    }
  ],
  "totalRows": 8,
  "totalColumns": 3,
  "_id": "8899cb23fc2642059c5bd08c8c1292cf"
}
```

Appendix 2

The descriptors calculated by the two tools, NanoXtract from NovaMechanics and NanoImage by NTUA, for circular/cylindrical/plates/other and Spherical nanomaterials respectively, are described in this Appendix, as a direct comparison of whcih tool might be used for which descriptors or User requirements, and to demonstrate where potential cross-checking of outputs could be tested by comparison of the two tools performance on a common set of images. This will be undertaken internally by NanoCommons partners as a next step in the validation of the tools.

First, a brief table is provided depicting the descriptors, a short description of each descriptor, the tools that produce them, and their values and units. Secondly, the explanation of descriptors follows. In Appendix 4 the ontological terms for all descriptors are provided.

The columns in the following table show:

- the descriptor name
- a brief explanation (a more detailed description follows the table)
- the tools within the NanoCommons infrastructure that provide calculations for the respective descriptor. In the case that different names are used within the tools, they are shown in the respective column
- the range of descriptor values and the unit of measurement, where applicable

Image descriptor	Brief meaning	NanoImage NTUA	NanoXtract NM	Range of values
Angle	The angle between the primary axis of the best fitting ellipse and a line parallel to the x-axis of the image.			0-180 [deg]
Aspect Ratio	The quotient of the division: *major_axis/minor_axis*.			>0 unitless
Boundary size	Total length of the NP's boundary.			>0 [nm]
Boxivity	The extent to which a NP approaches a rectangle.			0-1 [unitless]
Circularity	The degree to which a NP approaches a perfect circle.			0-1 [unitless]
Convexity	The NP's edge roughness.			0-1 [unitless]
Diameter	The NP's diameter.	Surface diameter	Diameter	>0 [nm]
Eccentricity	The measure of how much the NP deviates from being circular.			0-1 [unitless]
Extent	Alternative calculation of boxivity.			0-1 [unitless]

Feret Angle	The angle between the (min) Feret diameter and a line parallel to the x-axis of the image.			0-180 [deg]
Grey deviation	The deviation of the grey values used to count the mean grey value.			>0
Integrated Density	The product of Area and Mean Gray Value.			>0
Kurtosis	The fourth order moment about the mean.			unitless
Main elongation	The lengthening of the NP.			0-1 [unitless]
Major axis/ Major	The longest diameter of the best fitting ellipse to the NP.	Major	Major axis	>0 [nm]
Max Grey Value	The highest grey scale values appearing in the selection.			>0
Maximum Feret's diameter	The longest distance between any two points along the selection boundary.			>0 [nm]
Mean Grey Value	The average grey value within the selection.			>0
Min Grey Value	The lowest grey scale values appearing in the selection.			>0
Minimum Feret's diameter/ Min Feret	The shortest distance between any two points along the selection boundary.	Min Feret	Minimum Feret's diameter	>0 [nm]
Minor axis/ Minor	The shortest diameter of the best fitting ellipse to the NP.	Minor	Minor axis	>0 [nm]
Modal Grey Value	Stands for the most frequently occurring grey value within the selection.			>0
Perimeter	Total length of the NP's boundary.			>0 [nm]
Porosity	Provides a measure of porosity.			>0
Roundness	Compares the surface of the NP to the surface of the disc of diameter equal to major axis.			0-1 [unitless]
Size/Area	The area of the NP.	Area	Size	>0 [nm2]
Skeweness	The third order moment about the mean.			unitless
Solidity	The degree of the overall concavity or convexity of a NP.			0-1 [unitless]
Sphericity	The ratio of the surface area of a sphere to the surface area of the particle.			>0 unitless
Surface diameter	The diameter of a sphere having the same surface area as the projected particle.			>0 [nm]
Volume	The sum of all non-zero pixels episodes,			>0 [nm3]

	combining measures of contour elements.			
Volume to Surface	The diameter of a sphere having the same volume to surface ratio as the projected particle.			>0 [nm]

Table A1: Image descriptors extracted by NanoImage and NanoXtract

Explanation of descriptors for spherical compounds

Angle

The angle between the primary axis and a line parallel to the x-axis of the image, in the range of 0-180 degrees.

Aspect Ratio

The quotient of the division: major_axis/minor_axis.

Boundary size/Perimeter

The perimeter of the particle is defined as the to-tal length of the particle boundary. The image tool produces two values expressing the perimeter (Perimeter and Boundary Size) that may be different due to use of different calculation methods (Mikolajczyk et al., 2015).

Boxivity

The boxivity ratio expresses the extent to which a particle approaches a rectangle. It is defined as the actual area of the particle (A) divided by the area of the smallest enclosing rectangle (AR). The smallest enclosing rectangle (or minimum bounding box) is the smallest rectangle that contains every point in the particle (Toth et al., 2013). Another method is used to calculate this measure, producing the **Extent** descriptor (see below).

$$Boxivity = \frac{A}{A_R}$$

Circularity

The circularity ratio expresses the degree to which a particle approaches a perfect circle (value equal to 1.0). Both particle’s area and perimeter are taken into consideration, thus both form and roughness are measured. The more elongated becomes a shape, the more the circularity value approaches 0.0. In the same time the smoother a shape becomes, the more the length of the perimeter decreases, in comparison to the measured surface area (Mikolajczyk et al., 2015; Komsta, 2011; Ferreira, 2015). As a function of the surface area of the particle (A) and the particle's perimeter it is calculated using the following relation:

$$f_{circ} = \frac{4\pi A}{L^2}$$

Convexity

By calculating the convexity of a particle, its' edge roughness is defined. The actual perimeter (P) of a particle is defined as the exact length of its' boundary, whereas the convex hull perimeter (Pc) is the perimeter of the corresponding convex shape. As a shape becomes rougher (increase of the number and size of irregularities of the edges) the perimeter of the shape may increase rapidly. On the other hand, by surrounding the shape, the perimeter is smoothed into the convex hull, thus as the shape becomes rougher, the convex hull perimeter changes more slowly, and eventually approaches a large –but smooth- circle. The convexity of a smooth shape approaches 1.0 (Toth et al.; 2013, Komsta, 2011; Mingqiang et al., 2008).

$$Convexity = \frac{P_c}{P}$$

Diameter

The diameter of a spherical particle is any straight-line segment that passes through the center of the particle and whose endpoints lie on the boundary. For particles of different shape, the circular equivalent diameter, or area-equivalent diameter, is defined as the diameter of a circle with the same area as the particle (Mingqiang et al., 2008).

Eccentricity

The eccentricity is a measure of how much the particle deviates from being circular. As in the case of an ellipse, the two semi-major (a) axes and the semi-minor axis (b) are used (see major/minor axes below) (ImageJ, 2018).[7]

$$Eccentricity = \sqrt{1 - \left(\frac{b}{a}\right)^2}$$

Extent

Alternative calculation of boxivity. The boxivity ratio expresses the extent to which a particle approaches a rectangle. It is defined as the actual area of the particle divided by the area of the smallest enclosing rectangle. The smallest enclosing rectangle (or minimum bounding box) is the smallest rectangle that contains every point in the particle. Two values of this measure (**Boxivity** and **Extent**) are calculated using different methods (Toth et al., 2013; Komsta, 2011).

Feret Angle

FeretAngle (0-180 degrees) is the angle between the Feret diameter and a line parallel to the x-axis of the image.

Grey deviation

The deviation of the grey values used to count the mean grey value.

Integrated Density

Calculates and displays "IntDen", the product of Area and Mean Grey Value.

Kurtosis

The fourth order moment about the mean, providing a measure of flatness. A value equal to zero denotes the Gaussian (normal) distribution. Positive values are an indication of a distribution more peaked than normal. Negative values mean a distribution that is flatter than normal, however if less than -1.2, they denote a bimodal or multimodal distribution (Richard, 2007).

Main elongation

The elongation is calculated using the parameters of the minimum bounding box; the largest (L) and the shortest (S) sides of the minimum bounding rectangle, and expresses the lengthening of the particle (Odziomek et al., 2017; Olson, 2011).

$$Main\ elongation = 1 - \frac{S}{L}$$

Major axis/Minor axis

In order to calculate the major axis, a technique is applied to each particle that fits an ellipse in the optimal way in order to characterize in a quantitative manner the particles by a series of terms. The best-fitting ellipse has the same area as the initial particle and the measured maximum diameters of this ellipse correspond to the major axes respectively (Komsta, 2011; ImageJ, 2018).[4,7]

Max Grey Value

The highest grey scale value appearing in the selection.

Maximum Feret's diameter

The maximum Feret's (or calliper) diameters is defined as the longest distance, respectively, between any two parallel tangents on the boundary of a particle (Mikolajczyk et al., 2015; Mingqiang et al., 2008).

Mean Grey Value

Average grey value within the selection. In this value the sum of all pixels within the selection is divided by the count of pixels. If the image is not in grey scale then a transformation is made according to the following formula: *grey=0.299red+0.587green+0.114blue*.

Min Grey Value

The lowest grey scale value appearing in the selection.

Minimum Feret's diameter

The minimum Feret's (or calliper) diameters is defined as the shortest distance, respectively, between any two parallel tangents on the boundary of a particle (Mikolajczyk et al., 2015; Mingqiang et al., 2008).

Minor axis/ Minor

In order to calculate the minor axis, a technique is applied to each particle that fits an ellipse in the optimal way in order to characterize in a quantitative manner the particles by a series of terms. The best-fitting ellipse has the same area as the initial particle and the measured minimum diameters of this ellipse correspond to the minor axis respectively (Komsta, 2011; ImageJ, 2018).

Modal Grey Value

The most frequently occurring grey value within the selection.

Perimeter

Total length of the NP's boundary.

Porosity

Provides a measure of porosity, being equal to the feature Area Fraction calculated by ImageJ. Area Fraction is defined as the percentage of pixels in the image or selection that have been highlighted in red using the Threshold function. Users control the Threshold function by selecting one of the available filter types. For non-thresholded images, the percentage of non-zero pixels.

Roundness

Roundness is a measure that compares the surface area of a particle to the surface of a disc of diameter equal to its major axis (Mikolajczyk et al., 2015; Komsta, 2011).

$$Roundness = \frac{4A}{\pi(major\ axis)^2}$$

Size/Area

The size measures the surface area enclosed by the outer contour of the particle (Mikolajczyk et al., 2015).

Skeweness

The third order moment about the mean, providing a measure of symmetry. Zero values denote a symmetric distribution, negative values indicate a distribution asymmetric to the left, positive values a distribution asymmetric to the right (Richard, 2007).

Solidity

Solidity expresses the degree of overall concavity or convexity of a shape (the particle's area -A, divided by the convex hull area -Ac). As a particle becomes more convex/compact, both areas approach each other. The rates at which both areas change differ significantly, as the enveloping of the shape in a convex hull has the net effect of smoothing the shape, and finally the solidity value approaches 1.0. On the contrary as a particle's form deviates from a convex shape, the convex hull area increases and the solidity value decreases (Mikolajczyk et al.; 2015, Toth et al., 2013; Mingqiang et al., 2008). Defined as the quotient of: area/convex area.

$$Solidity = \frac{A}{A_c}$$

Sphericity

The ratio of the surface area of a sphere - with the same volume as the particle considered - to the surface area of the particle.

$$\psi = \frac{\pi^{\frac{1}{3}} 6V^{\frac{2}{3}}}{A}$$

Surface diameter

The diameter of a sphere having the same surface area as the projected particle.

$$d_s = \sqrt{\frac{A}{\pi}}$$

Volume

The sum of all non-zero pixels episodes, combining measures of contour elements, assuming that the contour is a square with a side equal to the unity.

Volume to Surface

Volume to Surface (Equivalent volume/surface) is the diameter of a sphere having the same volume to surface ratio as the projected particle.

$$d_{Sauter} = \frac{6V}{A}$$

Appendix 3

The descriptors calculated by the NanoImage tool that is provided by NTUA, for CNTs are described in this Appendix. The equivalent ontological terms for all descriptors are provided in Appendix 4.

The results for CNTs are separated into measurement identification values and descriptors. They are listed below, together with the Ridge Detection plugin term they correspond to.

CNT Measurement identification values

Pos

The numbering of each point in the CNT (defined as *num* in Ridge Detection), which equates to the length of a specific tube.

ID/Contour ID

Refers to the numbering of the CNT the point in column Pos belongs to. Its value is assigned in the plugin.

X coordinates / Y coordinates

The coordinates of the line points are given in the columns **X coordinates** and **Y coordinates** (named arrays *row* and *col* internally in the plugin).

CNT descriptors

Angle of Normal

The direction of the normal to each line point, as measured from the row-axis (array *angle* in Ridge Detection). The normal line to a curve at each line point is the line through that point and perpendicular to the tangent (see figure 5b of (Steger, 1998)). Some people like to call the col-axis the x-axis and the row-axis the y-axis, and measure the angle from the x-axis. Provides a measure of the straightness or curviness of the nanotube.

Contour Class

The contour class can assume values that characterise the the CNT line (lines are called contours in the plugin). An example of detection of junctions recognised in images of CNTs is provided by the author of the Ridge Detection plugin (Steger, 1998), showing the junctions recognised in lines that are connected. Please refer to Figure 5 of the publication. In this implementation of the plugin, CNTs are recognised as lines by the algorithm. Possible values of the contour class are:

- *cont_no_junc*: no end point is a junction
- *cont_start_junc*: only the start point of the line is a junction
- *cont_end_junc*: only the end point of the line is a junction
- *cont_both_junc*: both end points of the line are junctions
- *cont_closed*: the contour is closed

The values can indicate the morphology of the CNTs, for example how intertwined they appear to be in the microscope image.

Contrast and Asymmetry

The descriptors **Contrast** and **Asymmetry** (*contrast* and *asymmetry* in the plugin) contain the true contrast and asymmetry of each line point if the algorithm was instructed to apply the width and position correction. Otherwise, they are set to NULL. If the asymmetry, i.e., the weaker gradient, is on the right side of the line, the asymmetry is set to a positive value, while if it is on the left side it is set to a negative value.

Line Width (refers to CNT section) / **Mean Line Width** (refers to entire CNT)

Line Width equals to the sum (WidthL +WidthR), which contain the width information for each line point if the algorithm was requested to extract it; otherwise they are NULL (*WidthL* and *WidthR* in the plugin). If the line position and width correction was applied the contents of WidthLwidth_l and WidthRwidth_r will be identical. **Mean Line Width** equals to the mean value of this feature along the CNT.

Estimated Length (refers to CNT section) / **Length** (refers to entire CNT)

These descriptors provide the length either of the CNT section, or the entire CNT, respectively.

Appendix 4

This appendix provides the ontological terms of spherical NM image descriptors in the anoMapper ontology. It also indicates descriptors which are not currently supported by an ontological term.

In some cases, especially for CNTs, there are no terms available at the moment. NanoCommons will address this need within the course of the project and add these terms to the NanoCommons update of the eNanoMapper ontology, making them available.

Please note that all terms' definitions are provided in Appendix 2.

Spherical nanoparticles

Image descriptor	Ontology
Angle	No ontological term is available.
Aspect Ratio	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000064
Boundary size	No ontological term is available.
Boxivity	No ontological term is available.
Circularity	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000038
Convexity	No ontological term is available.
Diameter/Surface diameter	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000058
Eccentricity	No ontological term is available.
Extent	No ontological term is available.
Feret Angle	No ontological term is available.
Integrated Density	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000042
Kurtosis	http://bioportal.bioontology.org/ontologies/ENM?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000043
Main elongation	No ontological term is available.
Major axis/ Major	No ontological term is available.
Max Grey Value	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000047
Maximum Feret's	No ontological term is available.

diameter	
Mean Grey Value	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000044
Min Grey Value	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000046
Minimum Feret's diameter/ Min Feret	No ontological term is available.
Minor axis/ Minor	No ontological term is available.
Modal Grey Value	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000048
Perimeter	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FPATO_0001711&jump_to_nav=true
Porosity	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000051
Roundness	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000053
Size/Area	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000059
Skeweness	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000055
Solidity	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000056
Sphericity	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000057
Grey deviation	No ontological term is available.
Volume	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000061
Volume to Surface	http://bioportal.bioontology.org/ontologies/ENM/?p=classes&conceptid=http%3A%2F%2Fpurl.enanomapper.org%2Fonto%2FENM_8000063

Table A2 : Ontological terms for spherical NM image descriptors

Carbon nanotubes

Image descriptor	Ontology
Angle of Normal	No ontological term is available.
Asymmetry	No ontological term is available.
Contour Class	No ontological term is available.
Contrast	No ontological term is available.
(Estimated) Length	As they refer to the same physical attribute, (Estimated) Length refers to CNT section, while Length refers to entire CNT), they can be described by the same term, **Length**. However, no suitable term is available for Length of CNTs.
Line Width	As they refer to the same physical attribute (Line Width refers to CNT section, while Mean Line Width refers to entire CNT), they can be described by the same Ontology term, **Line Width**. No ontological term is available.

Table A3 : Ontological terms for CNT image descriptors