DELIVERABLE REPORT D5.5

User application for searching and downloading eNanoMapper data

GRANT AGREEMENT:	604134
ACRONYM:	eNanoMapper
NAME:	eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment
PROJECT COORDINATOR:	Douglas Connect GmbH
START DATE OF PROJECT; DURATION:	1 February 2014; 36 months
PARTNER(s) RESPONSIBLE FOR THIS DELIVERABLE:	IDEA
DATE:	20.04.2016
VERSION:	1.1

FP7-NMP-2013-SMALL-7

Document Type	Deliverable Report
WP/Task	WP5
Document ID	eNanoMapper D5.5
Status	Final Version

Partner Organisations	● Douglas Connect, GmbH (DC) ● National Technical University of Athens (NTUA) ● In Silico Toxicology (IST) ● Ideaconsult (IDEA) ● Karolinska Institutet (KI) ● VTT Technical Research Centre of Finland (VTT) ● European Bioinformatics Institute (EMBL-EBI) ● Maastricht University (UM) ● Misvik Biology (MB)
Authors	Nina Jeliazkova (IDEA) Nikolay Kochev (IDEA) Vedrin Jeliazkov (IDEA) Egon Willighagen (UM) Friederike Ehrhart (UM) Final review and edit by Barry Hardy (DC)
Purpose of the Document	Report on searching data capabilities
Document History	1.Table of Contents, 11/12/2015 2.First draft, 14/12/2015 3.Second draft 15/02/2016 4.Completed version, 20/04/2016

TABLE OF CONTENTS

Table of Contents

GLOSSARY

Abbreviation / acronym	Description
Content type	A media type (also MIME type and content type) is a two-part identifier for file formats and format contents transmitted on the Internet https://en.wikipedia.org/wiki/Media_type
JSON	JavaScript Object Notation is an open-standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is the most common data format used for asynchronous browser/server communication.
JSON-LD	JavaScript Object Notation for Linked Data, is a method of encoding Linked Data using JSON https://en.wikipedia.org/wiki/JSON-LD
RDF/XML	RDF/XML is a syntax, defined by the W3C, to express an RDF graph as an XML document https://www.w3.org/TR/rdf-syntax-grammar/
N3	Notation3, or N3, is a shorthand non-XML serialization of Resource Description Framework models
MDL MOL file, SDF file	MDL Molfile is a format for holding information about the atoms, bonds, connectivity and coordinates of a molecule. "SDF" stands for structure-data file and may consist of multiple MDL Molfiles, representing multiple molecules
REST	Representational state transfer (REST) is an abstraction of the architecture of the World Wide Web; more precisely, REST is an architectural style consisting of a coordinated set of architectural constraints applied to components, connectors, and data elements, within a distributed hypermedia system. REST ignores the details of component implementation and protocol syntax in order to focus on the roles of components, the constraints upon their interaction with other components, and their interpretation of significant data elements.

1. EXECUTIVE SUMMARY

Searching and downloading nanomaterial data was identified from requirements analysis as one of the top priority use cases for users within the Nano Safety community. The search modes include both free text and structured search, which appeal to different types of users. The typical structured search requires combining queries on chemical composition and query on physicochemical characterization and bioassay outcomes. The preferred way of displaying the search result is a summary, with a subsequent drill down for details. This deliverable describes a user friendly free text search application (http://search.data.enanomapper.net), provides examples of using the API for searching the eNanoMapper database via R, and examples of building visual summaries with Javascript. Download of substances and studies as well as search results is enabled via REST API calls. The download formats supported include JSON, spreadsheets (XLSX) and semantic formats (RDF/XML, N3, JSON-LD). The web pages at http://data.enanomapper.net provide download links for user convenience. A blog post, describing how to use the different database search functionalities and an example of measurement representations in a semantic format are provided as Annexes.

2. INTRODUCTION

The requirements analysis performed by eNanoMapper within the scientific community (WP1) identified searching and downloading nanomaterial data as one of the top priority use cases for eNanoMapper (D1.2). Structured and free text search appeal to different types of users (persona). The typical structured search requires combining queries on chemical composition and query on physicochemical characterization and bioassay outcomes (described in eNanoMapper report on “Data Management System with extended search capabilities” (D3.2) and in the eNanoMapper database publication [1]). We have implemented a free text search application, which is powered by the Apache SOLR [2] free text search engine and is using the eNanoMapper ontology [3] for database entries annotation and synonym expansion. The REST API enables search through programming environment, e.g. R [4] and building visual summaries through JavaScript. The visual summaries can be integrated in the eNanoMapper web interface, as well as embedded in external web sites. Download of substances and studies as well as search results is enabled via REST API calls and specifying the content type; it can be used via both web browser and programming environments.

3. USER APPLICATION FOR SEARCHING AND DOWNLOADING ENANOMAPPER DATA

The eNanoMapper prototype database, available at http://data.enanomapper.net/ and described in eNanoMapper reports on the eNanoMapper database develoment (D3.1, D3.2) and in the eNanoMapper database publication [1], supports several query types and allows retrieving information of nanomaterials through a REST web services API (http://enanomapper.github.io/API/) and a web browser interface. This deliverable describes a user application providing an intuitive free text search interface as well as using the database API to send queries through R and JavaScript. Data download is supported via REST API and allows RDF, CSV, JSON and XLXS download.

3.1 SEARCH

The API offers access to a variety of searches by substance identifier, any combination of measurement endpoints, and/or chemical structure. The analysis of the use cases and the obtained feedback of using the prototype database highlighted the need of a simplified search interface, supporting synonyms and allowing many variants of specifying the query (e.g. ZnO, zinc oxide, zinc oxide nanoparticle, NPO_1584 return the same set of results). In addition, browsing the results can be facilitated by providing summary statistics of the elements found, e.g. the number of metal oxide entries. Therefore a free text search user interface was built, allowing data exploration without requiring *a priori* knowledge of the database content and field names. The free text search application uses the Apache Solr [2] search engine and the eNanoMapper ontology as synonyms and database entries annotation.

An initial version of the free search interface was developed as a single page JavaScript application and a screenshot submitted to the eNanoMapper issue tracker for feedback by partners and interested users https://github.com/enanomapper/data.enanomapper.net/issues/8 (Figure 1).

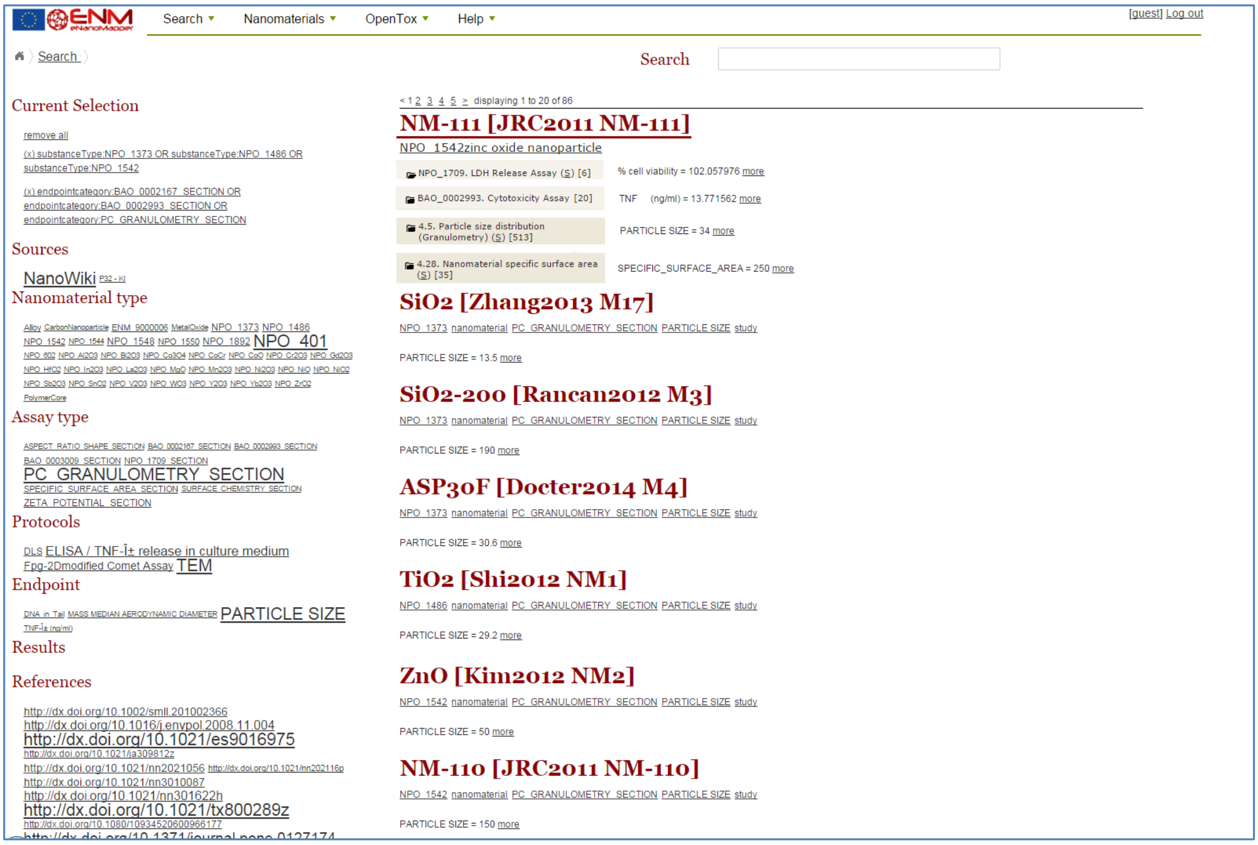


Figure 1. Initial interface of the free text search application

The feedback provided by all eNanoMapper partners was positive, with many specific suggestions, summarized below:

- *The display should use labels instead of NPO_xxx codes*
- *add the ability to "sort by", e.g. alphabetical on name, on size, year of the experiment*
- *the references could use the DOI without the http://dx.doi.org/*
- *having the possibility of multiple selection, getting the, listing identification, characterization, assay, results*
- *Search should be possible for e.g. 'ZnO' or 'zinc oxide', not only NPO code*
- *Searching by test species*
- *Search for specific guidelines*
- *The difference between Assay types and Endpoints is not immediately obvious*
- *The purpose of the result section is unclear*
- *Font size in the text clouds are confusing, they indicate frequency and not selected items*
- *A hierarchy would be useful e.g. for data aggregation (NM types and assays/endpoints). Low priority.*
- *Should there be a group for selecting e.g. particle size ranges? - Yes, but low priority.*

Taking into account the feedback, the application was further developed and made available at https://search.data.enanomapper.net and http://enanomapper.github.io/data.enanomapper.net/

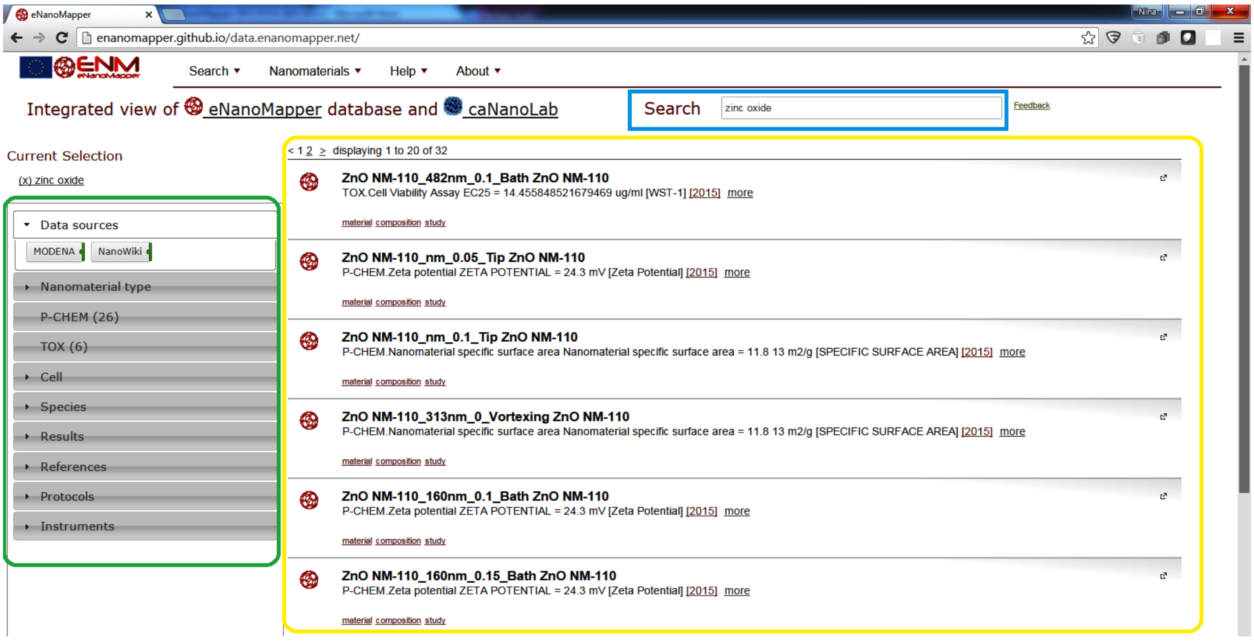


Figure 2. Search application (zinc oxide query)

The search application is a single page web application (Figure 2), using asynchronous JavaScript calls to submit queries, receive results and to render the web page elements. The page consists of a search box (highlighted in blue, top) a summary panel (left, green) and the results are shown at the main panel (yellow). The results of a search for “zinc oxide” are illustrated on Figure 2. The panel at the left (green) contains several summaries, grouped in the following categories: *Data sources, Nanomaterial type, P-Chem (physicochemical characterization), Tox (toxicity assays), Cell, Species, Results, References, Protocols, Instruments*. Every panel is expandable and shows the types of elements found for the particular query, *“zinc oxide”* in this case. For example, there are two data sources shown on Figure 2, “*MODENA*” and “*NanoWiki*”, because these two data sources contain entries for zinc oxide particles. The available data sources in http://data.enanomapper.net are described in eNanoMapper reports D3.2 and previously in eNanoMapper reports D3.1 and D5.4.

All widgets on the left are expandable on clicking. Figure 3 shows the expanded *NanoMaterial type* widget.

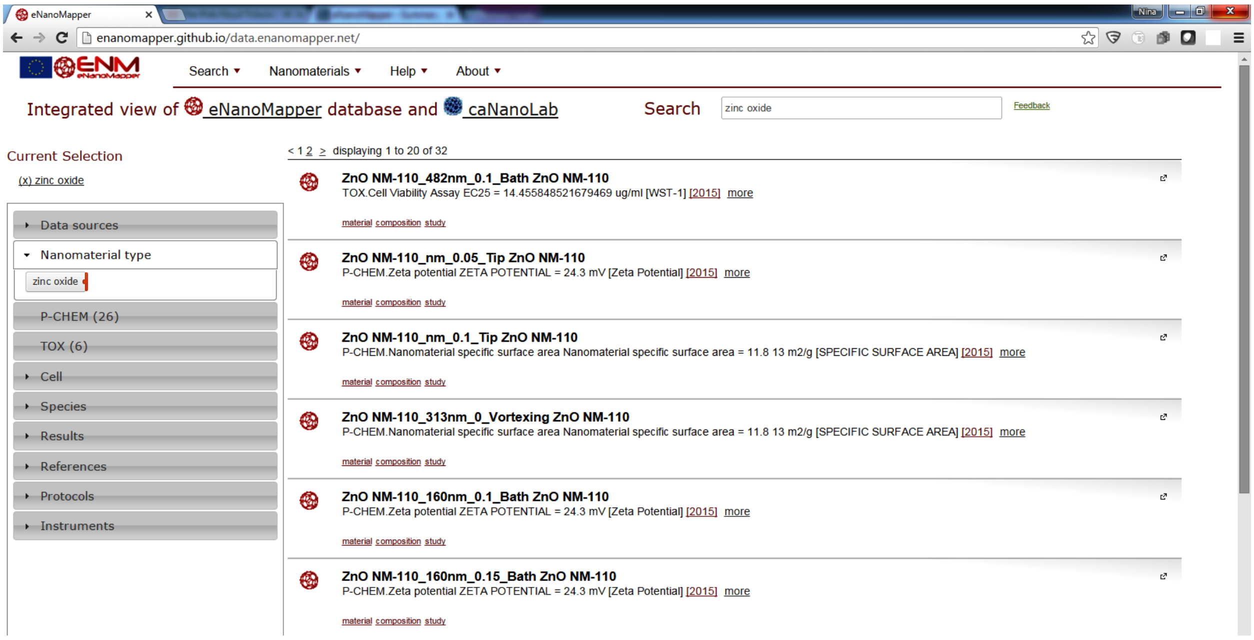


Figure 3. The nanomaterial type panel shows zinc oxide only, because this is what the query is about.

Figure 4 shows the expanded P-CHEM widget, where P-CHEM is short for physico-chemical characterisation and shows a summary of the type of experiments (the tags marked with green line at the right) and endpoints (the tags marked with blue line at the right).

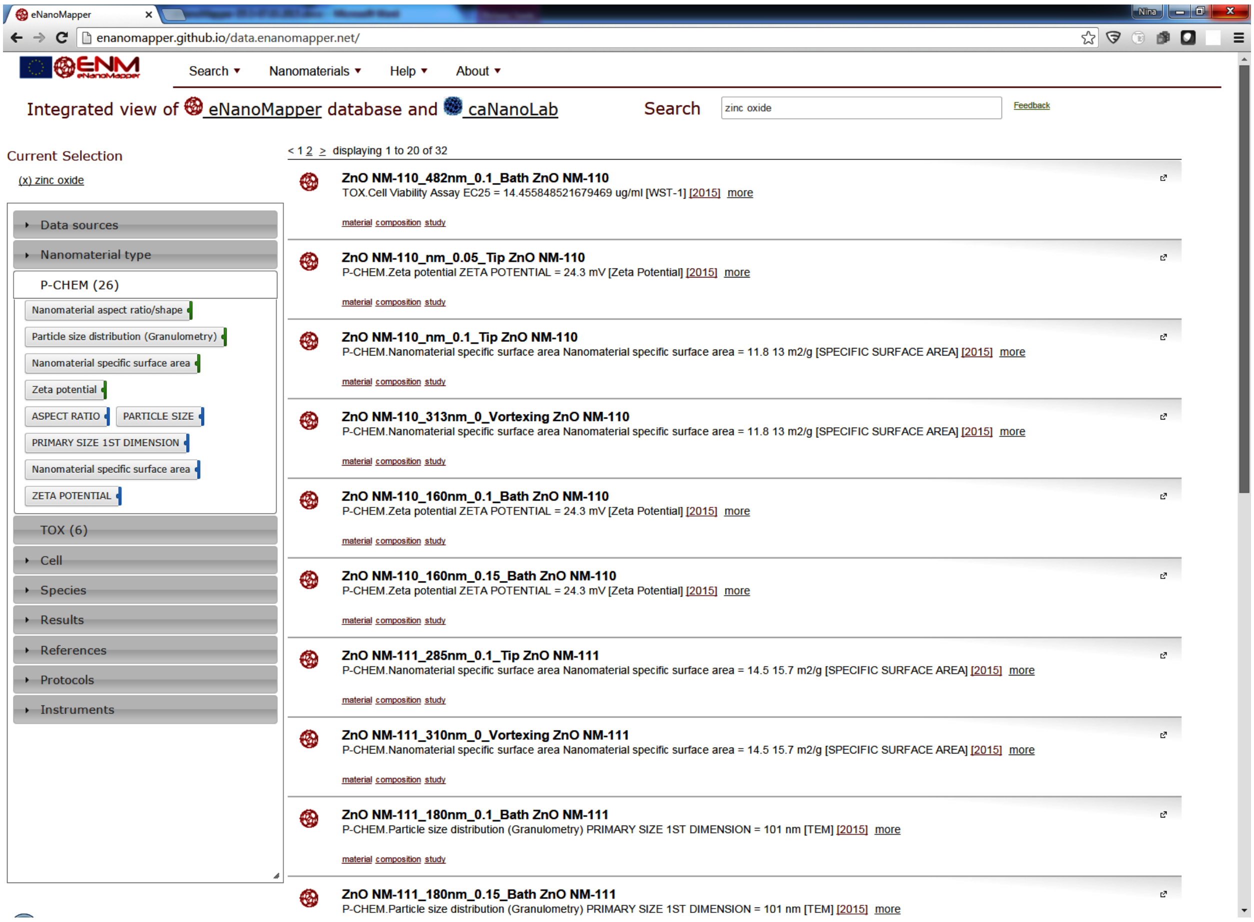


Figure 4. The P-CHEM (physico-chemical characterisation) panel shows the type and the number of entries available for different physicochemical measurements.

Mouse hover on each tag reveals more information, such as number of entries (the coloured part of the tag) or ranges of the available measurement (tooltip on the tags marked blue) - see Figure 5.

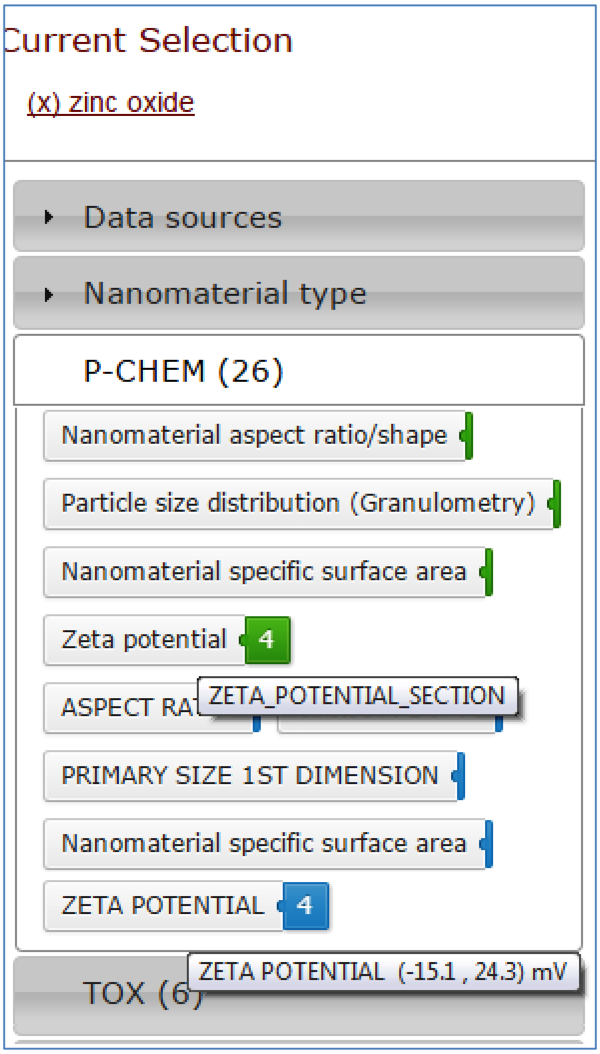


Figure 5. Summary details, e.g. shows the range of the Zeta potential measurements (-15.1, 24.3) mV

Figure 6 shows the *TOX* widget expanded. TOX stands for toxicity assays, and shows a summary of the type of the experiments (the tags marked with green line at the right) and endpoints (the tags marked with blue line at the right).

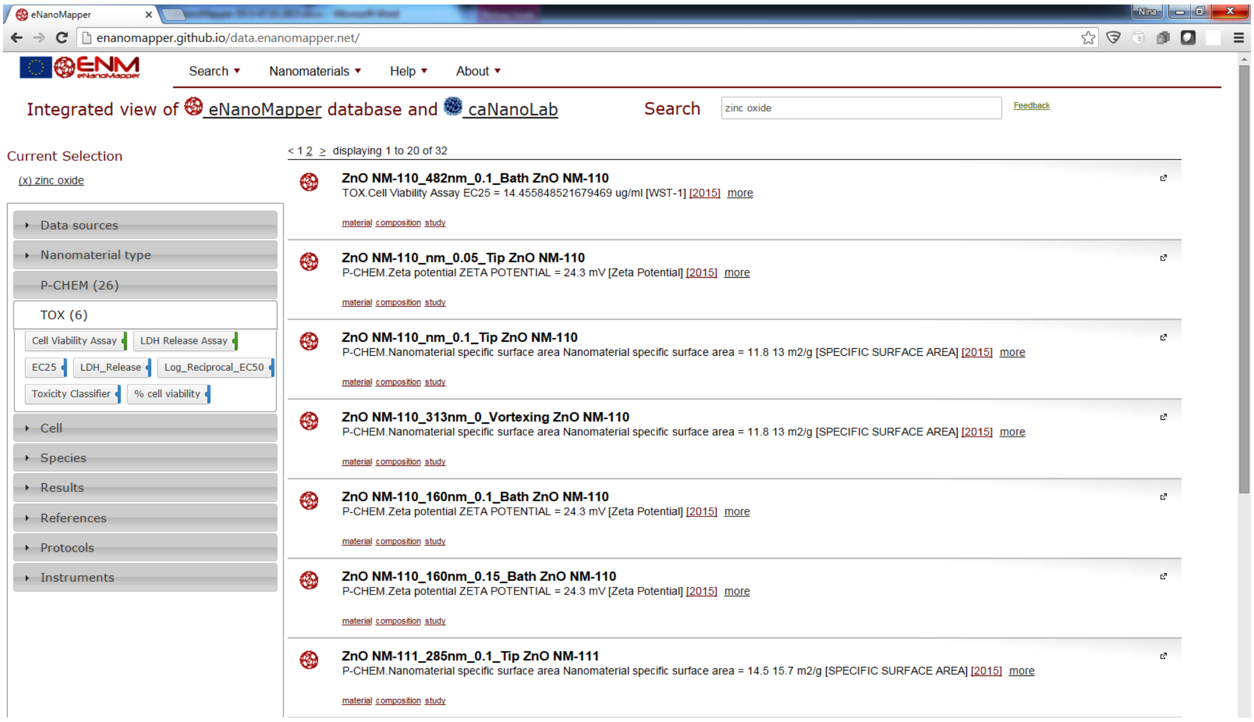


Figure 6. The TOX (toxicity assays) panel shows the type and the number of entries available for different biological assays. Hover with mouse on each tag in order to see more details

Figure 7 shows the *Cell widget* expanded. The *Cell widget* displays the cell lines used in the biological assays involved with zinc oxide (recall we are still exploring the “zinc oxide” search results, and this is shown under the *Current selection* section at the top left).

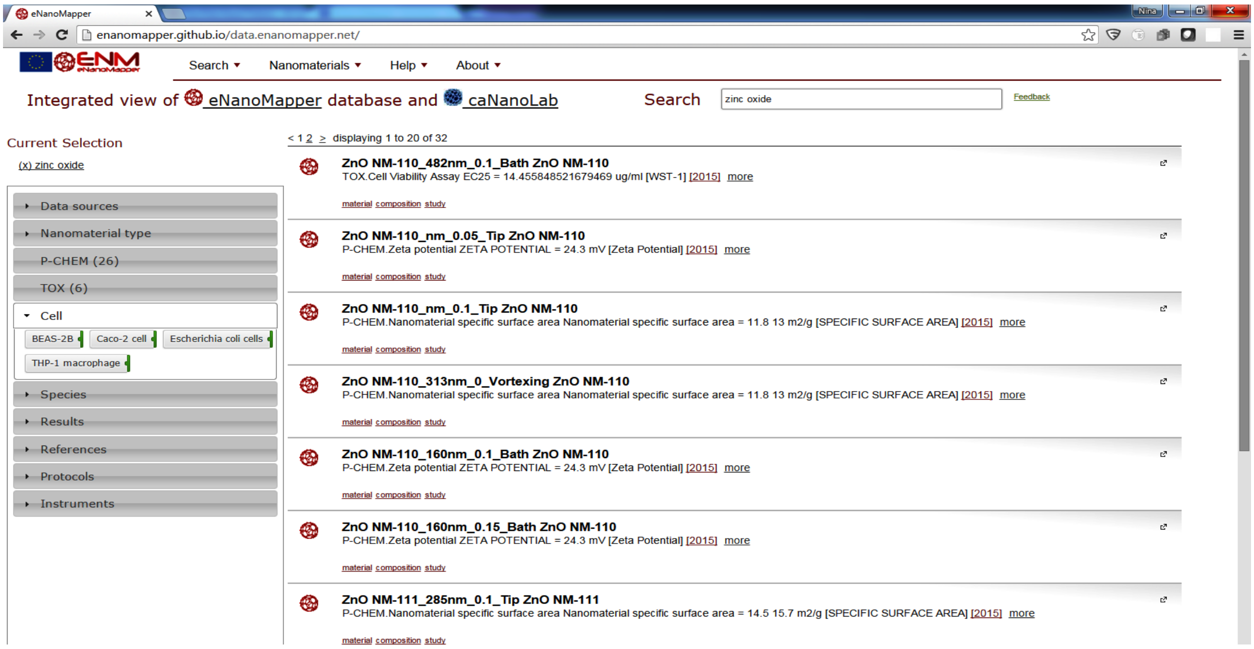


Figure 7. The cell lines used in the experiments, involving zinc oxide nanoparticles

Clicking on each button updates the Current selection with another entry. Clicking on the *THP-1 macrophages* button adds a second entry, restricting the query for *zinc oxide + THP-1 macrophage*. This is an illustration of how one can restrict the query by clicking any combination of the tags from the summary panels. The content of the result and summary panel will adapt to reflect the query results (Figure 8).

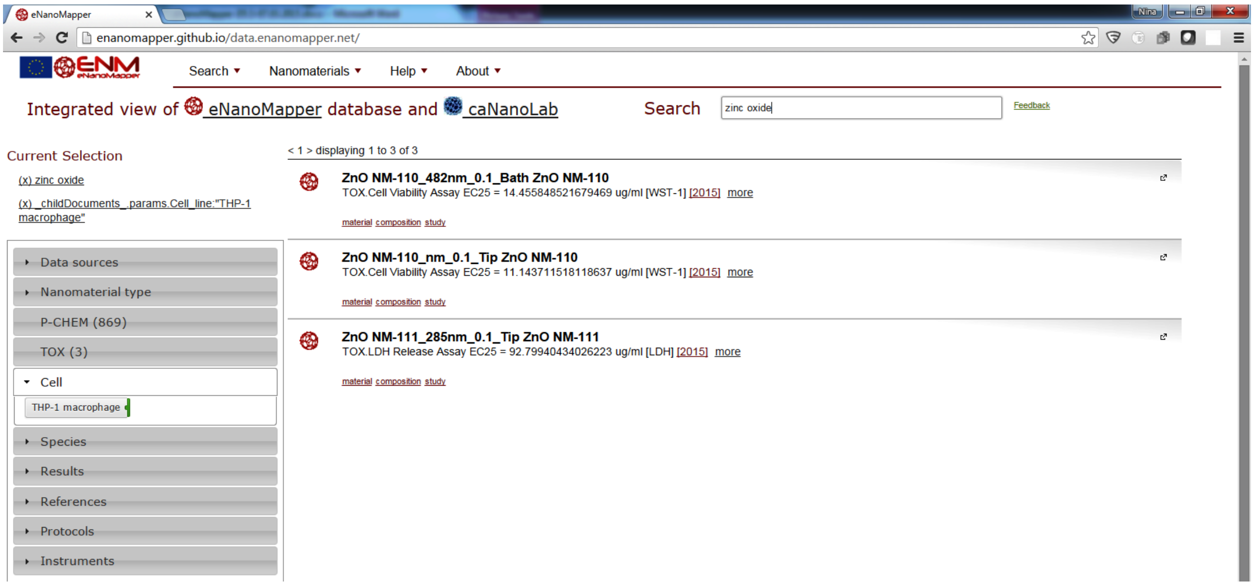


Figure 8. The cell lines used in the experiments, involving zinc oxide nanoparticles.

The content of the *Current selection* can be directly modified by clicking the (x) links, which will remove the selection and the result and summary panel content will be updated (Figure 9).

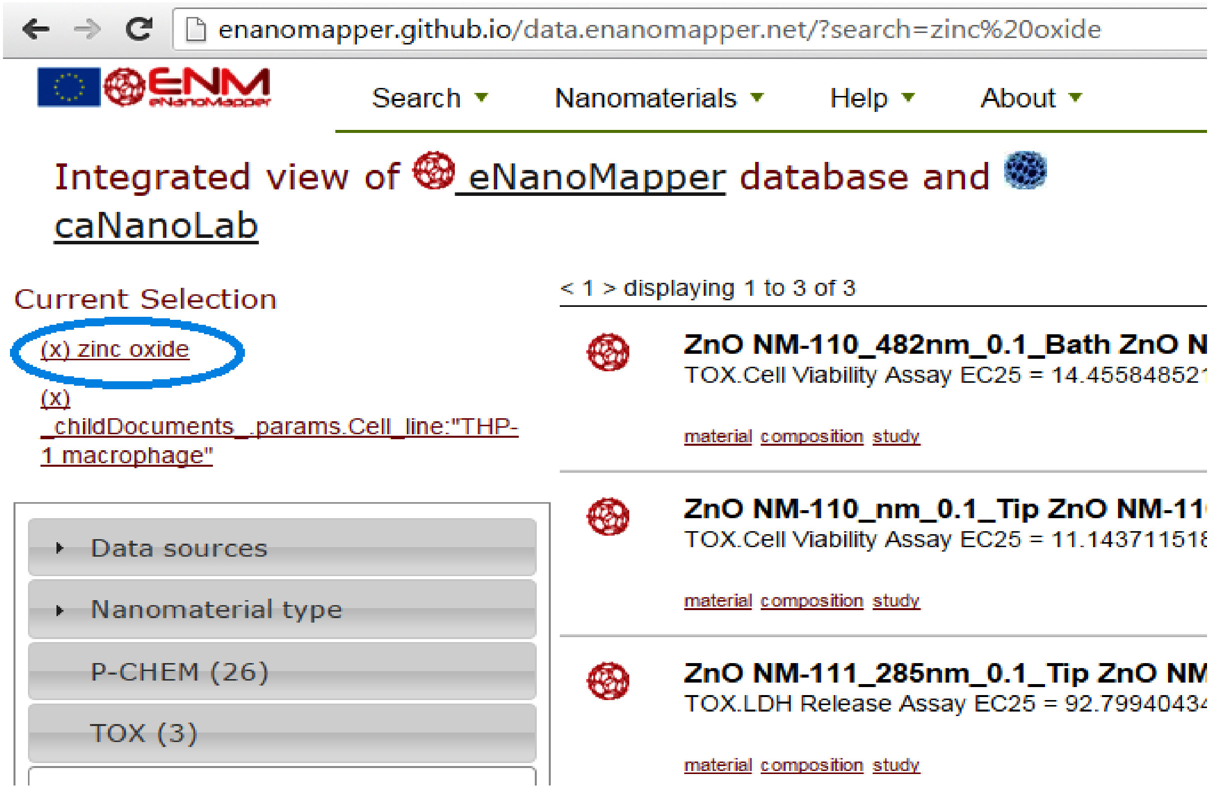


Figure 9. Click on the zinc oxide link to modify the current selection

As the results list is now updated, it contains particles other than zinc oxide (Figure 10).

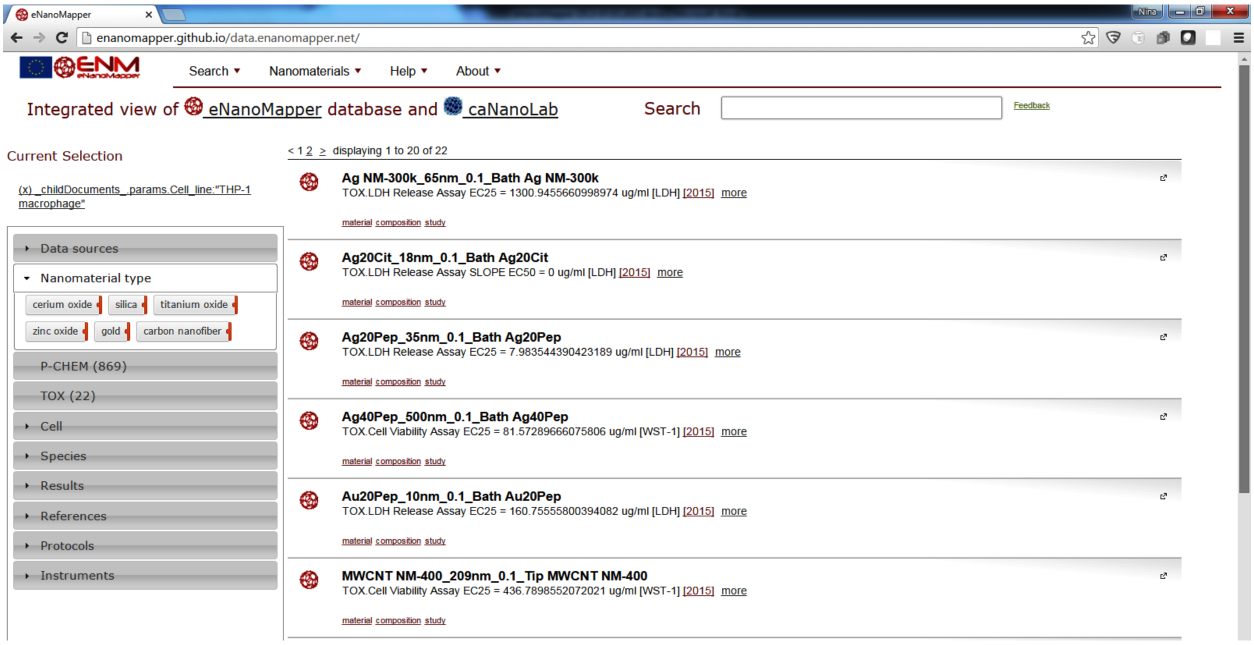


Figure 10, The expanded nanomaterial type widget on the left displays the summary of nanomaterial types found in the current query.

The links marked blue (Figure 11) lead to the full study record of the selected nanomaterial.

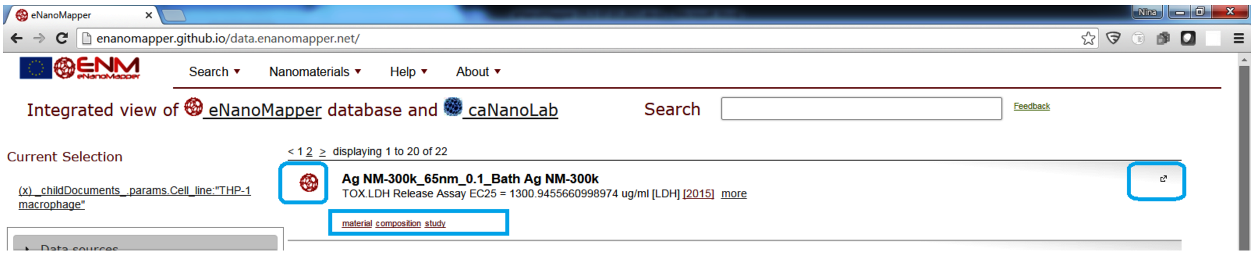


Figure 11. Links marked blue will lead to a database page exploring the full record of the nanoparticle characterisation.

The updated version resolves most of the issues, as follows:

- *NPO codes are resolved to ontology labels*
- *The refs use the DOI without the http://dx.doi.org/*
- *The free text search should behave user friendly, e.g. zinc oxide, zno, NPO_1542 should give the same result, also for any other phrase. This is achieved by using the eNanoMapper ontology (second release)[1] as list of synonyms*
- *The assay categories are split into phys chem and tox.*

[1] http://bioportal.bioontology.org/ontologies/ENM

- *The tag clouds are replaced by colour buttons*
- *The green tags within each category are assay types, the blue tags are endpoints.*
- *The colour part of the tag extends on hover and shows the number of entries having the tag. For endpoints there is a tooltip showing min/max values*
- *Searching for species and protocols is covered by the text search, but also there is a widget showing tags for protocols and species*
- *The appearance of all entries is uniform, there are links leading to the database entries*
- *Searching can be done by free text as well as ontology codes*
- *Hierarchical display will be developed in a future version (the query already returns hierarchical summaries)*

The link to the new search application is included in the main page of http://data.enanomapper.net

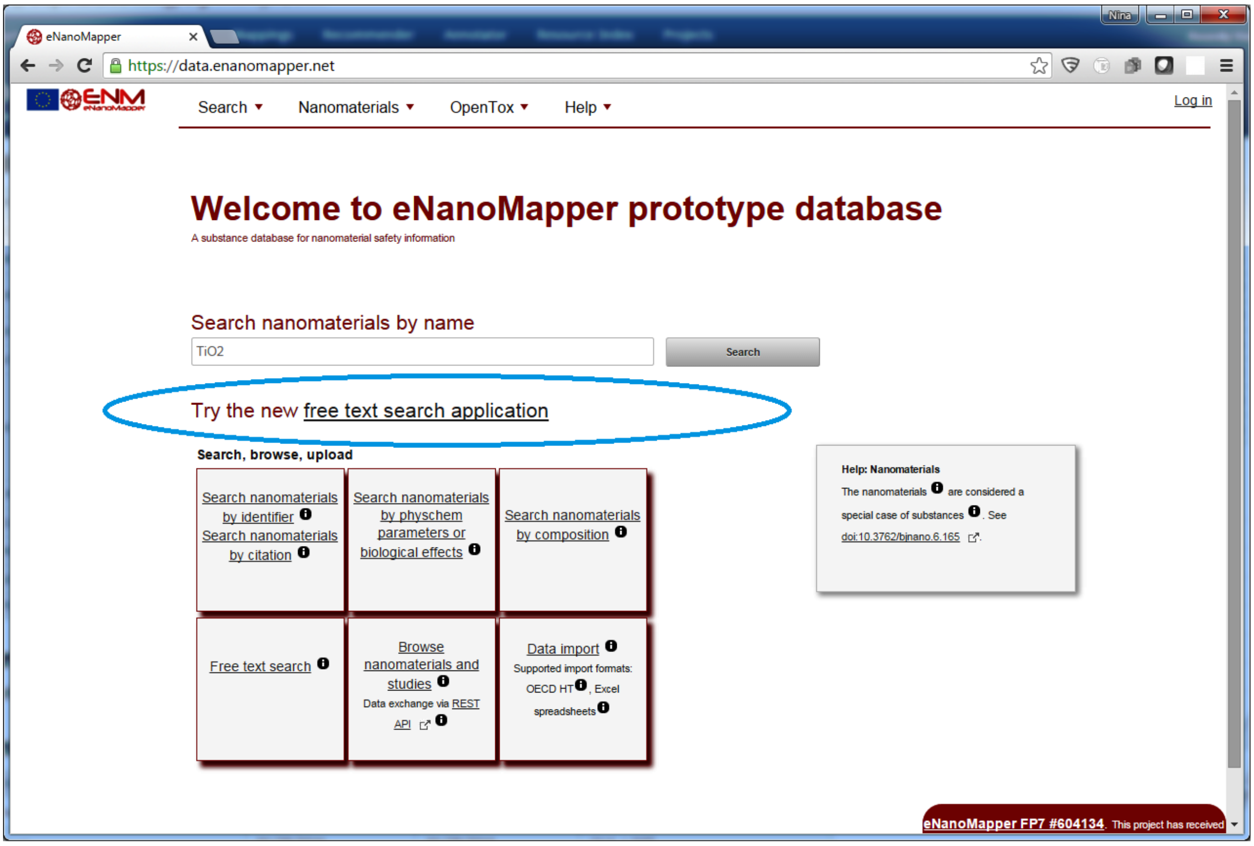


Figure 12. The main screen on data.enanomapper.net with a link to the free text search application.

3.2. SEARCH INTEGRATION

Besides providing a convenient interface for data exploration, the search application illustrates the integration between the eNanoMapper database and NCI caNanoLab [6], [7] database, online at https://cananolab.nci.nih.gov/caNanoLab/. The screenshot in Figure 13 shows that the results of searching for *"nanotube"* include both eNanoMapper entries and caNanoLab entries. This is achieved by configuring the search engine to build a common index for the content of both databases, while each of the databases is hosted separately. The content of the caNanoLab database was provided to eNanoMapper by caNanoLab developers as a MySQL dump.

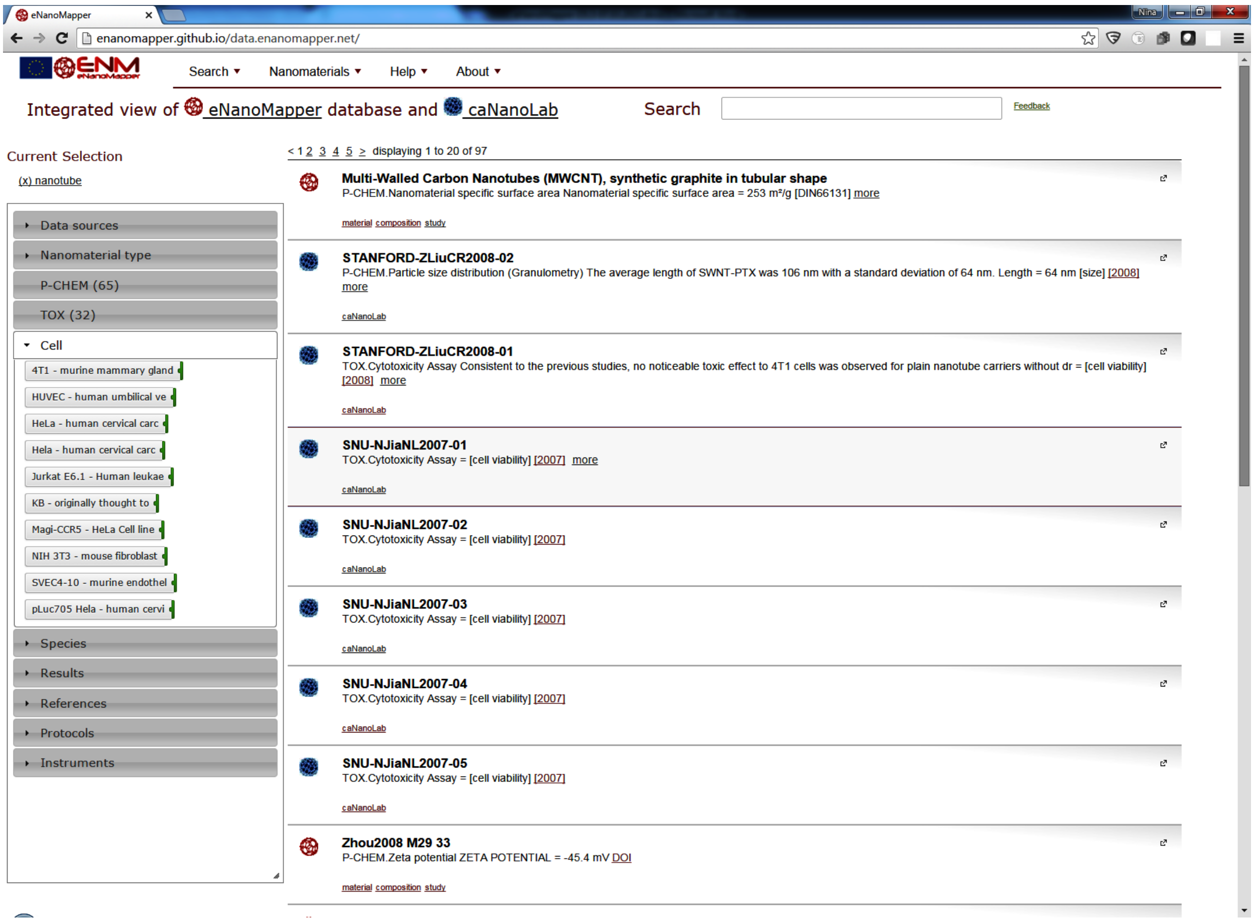


Figure 13. The screen shows results from both eNanoMapper database (red fullerene-like icon) and caNanoLab database (blue fullerene-like icon). Clicking on each icon will lead to the study records in eNanoMapper or caNanoLab respectively. This is a demonstration of virtual search integration between the two databases, physically residing in servers in Europe and US, but with a common search interface

3.3. JAVASCRIPT VISUAL SUMMARIES

Because the search functionality is also exposed via the API (see Figure 14), searches can also be done by API clients. An example API client is the JavaScript client (https://github.com/enanomapper/ambit.js/). Using this client library we can search for particular nanoparticles, such as titanium dioxide. To demonstrate searching via the API, a number of HTML5 web pages have been set up: http://enanomapper.github.io/enmSummaries/ . These pages summarize the information in data.enanomapper.net for silver, titanium dioxide, copper oxide, and zinc oxide. In four graphical plots and two tables, each page presents an overview of physical chemical and toxicology related properties (see Figure 14).

The JavaScript code to perform this kind of search follows quite closely the API itself. For example, the code used to search for titanium dioxide materials (note that the search results in Figure 14 also show materials outside the typical nano range) uses the API method to find substances related to the TiO2 chemical compound (which has the IRI https://data.enanomapper.net/compound/85/conformer/85):

```
var searcher = new Ambit.Substance("https://data.enanomapper.net/");
compound = "https://data.enanomapper.net/compound/85/conformer/85"
searcher.listForCompound(compound, processList);
```

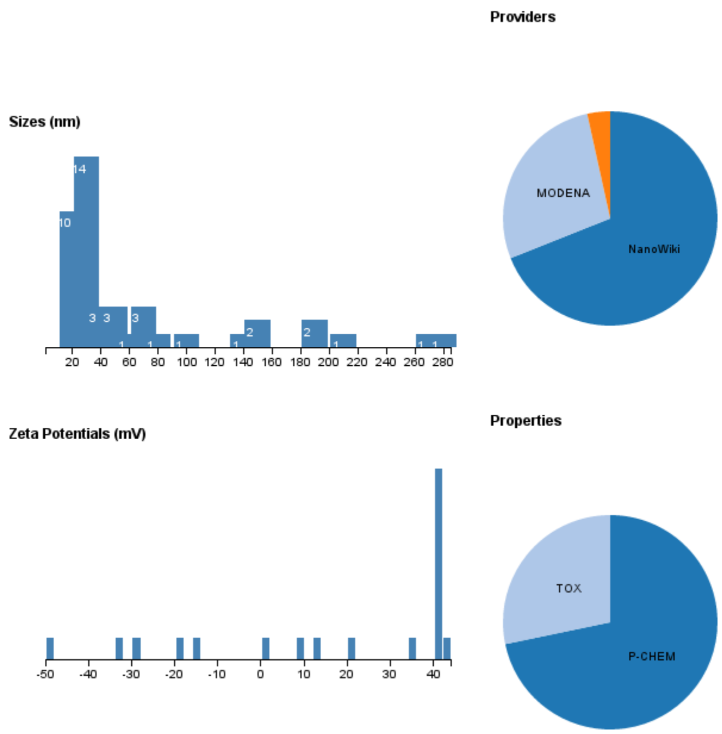


Figure 14. Screenshot of the graphical summary of available data in data.enanomapper.net of titanium dioxide.

However, other searches are possible too. For example, we can use the API method to search substances using the ambit.js wrapper *Ambit.Substance.search()*. This JavaScript method takes two parameters (besides the callback function), one for the search, the other for the type, which is the same type parameter in the API at http://enanomapper.github.io/API/#!/substance_1/getSubstances. This way, we can search for all materials which have been described in the Beilstein Journal of Nanotechnology [1]:

```
var searcher = new Ambit.Substance(
  https://apps.ideaconsult.net/enanomapper
);
searcher.search("10.3762/bjnano", "citation", processList);
```

3.4. SEARCHING IN R

A second client library that supports searching is the *renm* package for the R statistical environment. The approach of this client library is quite similar and closely follows the API. This R code example matches the ambit.js example for listing materials in some publication, though a specific PNAS paper here:

```
substances <- listSubstances(
  service="http://data.enanomapper.net/",
  search="10.1073/pnas.0802878", type="citation"
)
```

Another search example is the following code that lists all materials from a data set provider. Clearly, the diversity in the returned substances is enormous:

```
substances <- listSubstances(
  service="http://data.enanomapper.net/",
  search="NanoWiki", type="owner_name"
)
```

3.5. DOWNLOAD

The eNanoMapper database download facilities rely on the retrieval of resources through the defined REST API. Recalling that materials (substances), physicochemical characterization, bioassay data for materials, predefined sets of materials as well as search results are all REST resources, the API allows to export (an) arbitrary subset(s) of data in the supported formats. As REST allows for the provision of representations in multiple formats by specifying the content type of the HTTP call, the download links use the existing eNanoMapper API calls and specify the appropriate content type. The supported content types are listed in the table below.

Supported format	Content type
JSON	application/json
CSV	text/csv
XLSX	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
RDF/XML	application/rdf+xml
RDF N3	text/n3
JSON-LD	application/ld+json

An example of an API call to download material information in specified format by cURL:

```
curl "https://data.enanomapper.net/substance/XLSX-fa48e134-54cf-3417-a63a-6f744191cec2" -H "Content-type=?application/json"
```

An example of an API call to download material information in specified format, as a direct link:

https://data.enanomapper.net/substance/XLSX-fa48e134-54cf-3417-a63a-6f744191cec2?media=application/json

In addition to API calls, the eNanoMapper database web pages provide convenient links for download of resources in supported formats (see the JSON, CSV, XLSX, RDF and JSON-LD links in Figure 15).

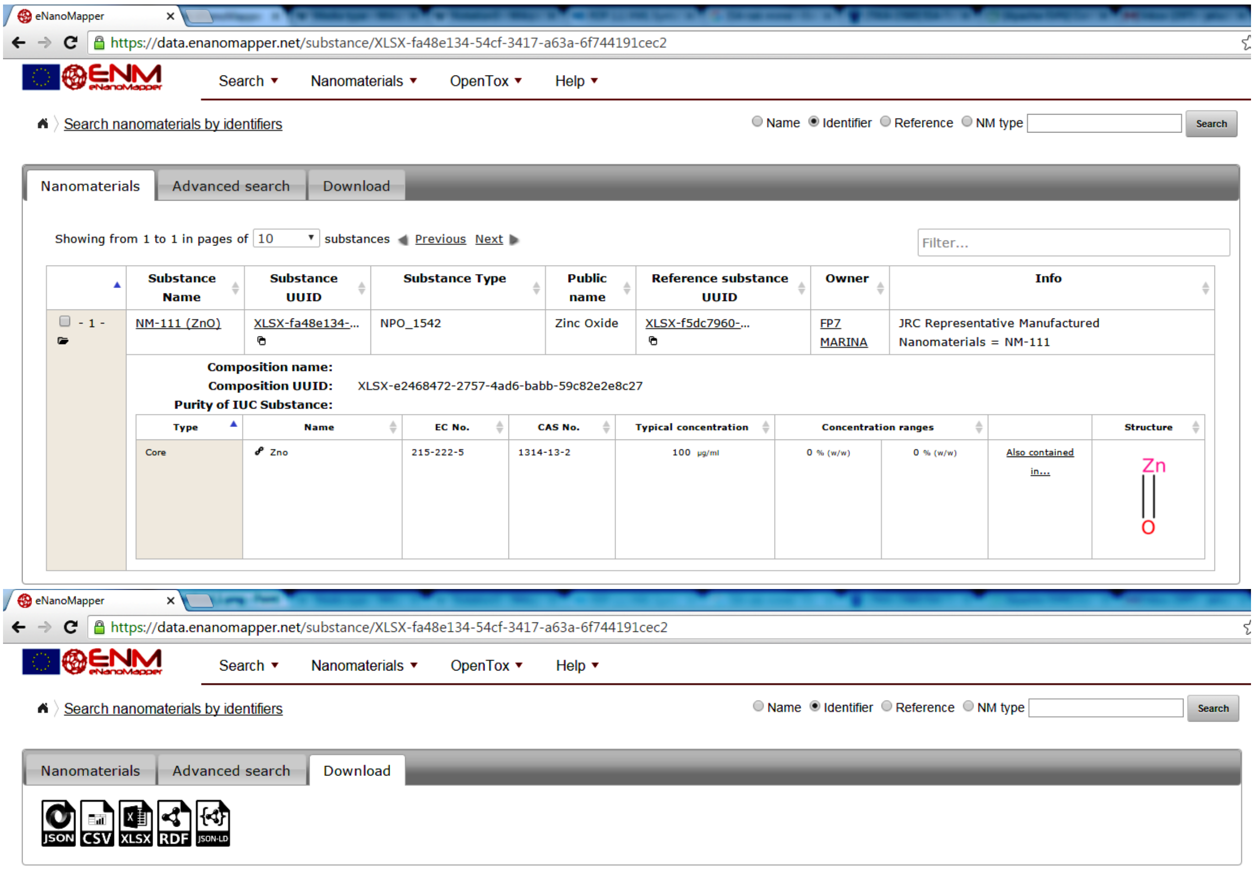


Figure 15. Substance and study download links for https://data.enanomapper.net/substance/XLSX-fa48e134-54cf-3417-a63a-6f744191cec2

The download links of search results are also available in the *Download* tab (Figure 16).

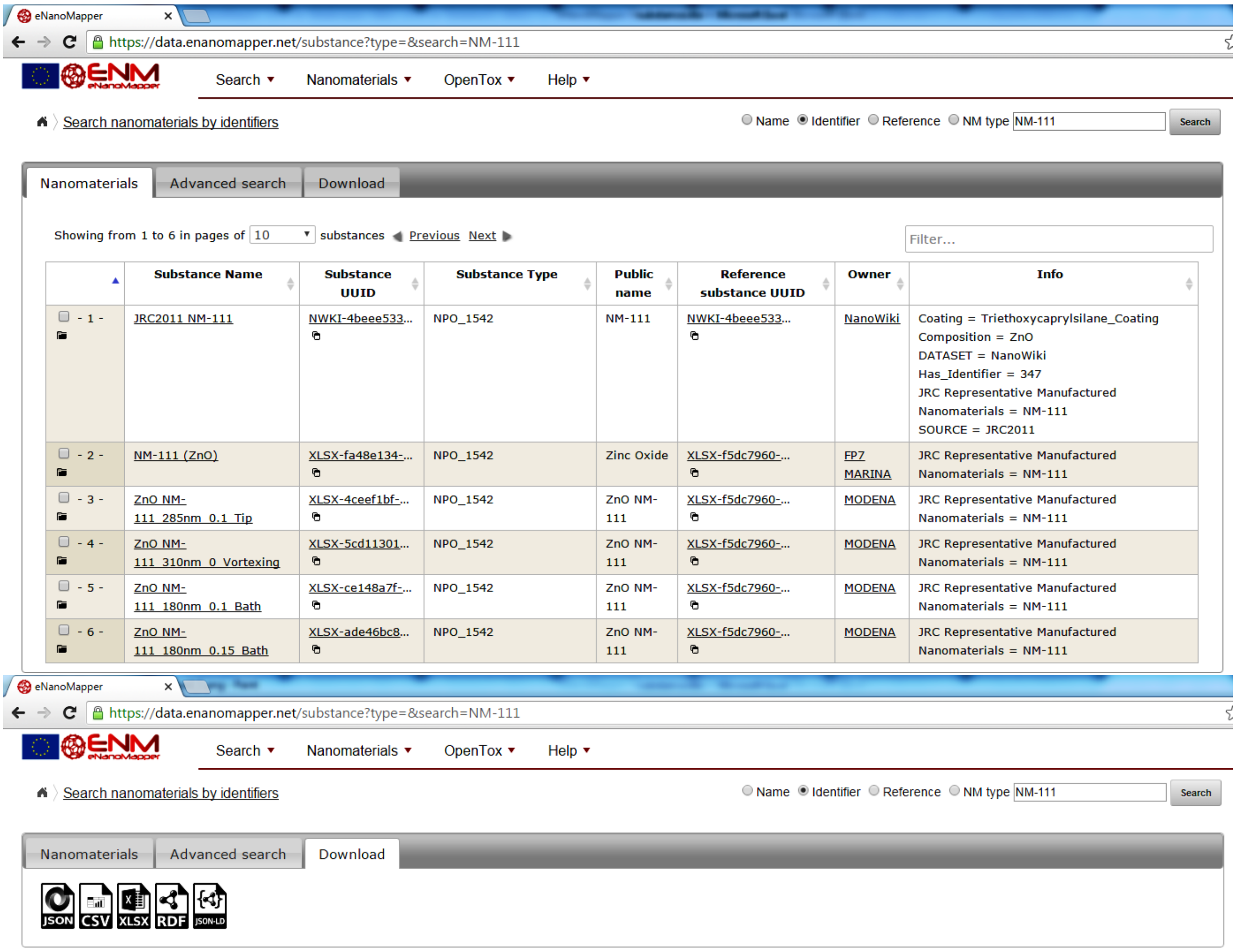


Figure 16. Download links for results of query for NM-111; https://data.enanomapper.net/substance? search=NM-111

The download links for a bundle are illustrated at Figure 17. A "bundle" is a REST resource that groups a selected set of substances and a selected set of endpoints [1]. This functionality enables creating groups of diverse nanomaterials, to specify the endpoints of interest, which can vary from physicochemical to proteomics assays, and to enable retrieving all this data with a single REST call. A bundle may include the nanomaterials and assay data from a single investigation as well as serve as a container for a set of NMs and for data (typically representing different experiments) retrieved from the literature. There are different download links, allowing exporting specific components of a bundle. The *substances* download link exports all substances (materials) from a bundle, the output formats and content are the same as described above (Figure 15 and Figure 16) for the substances and search results download. The *chemical structure* download links allows exporting all chemical structures, found as components of the materials in a bundle, the supported formats are as defined in the OpenTox Compound API (e.g. well known chemical formats as SDF). The *Properties* download link exports information about the measurements metadata in a format defined by OpenTox Feature API. The *Dataset* download links export the bundle substances and measurements in a format similar to the one defined by the OpenTox Dataset API, with additional export as Excel.

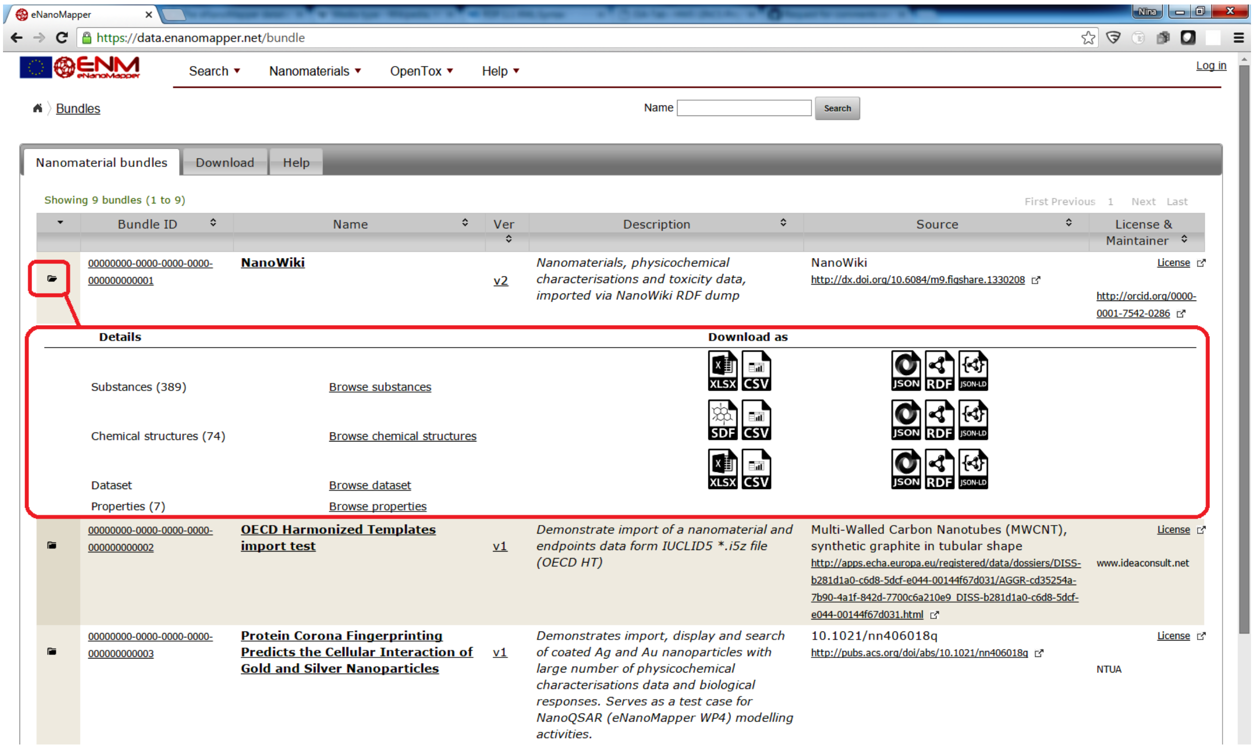


Figure 17. Download links for NanoWiki bundle, https://data.enanomapper.net/bundle .

The primary download format is JSON, representing the internal data model of materials and measurements, as defined in the API and described in [1], [8]. The RDF representation of the substances and experimental data follows the BioAssay Ontology (BAO) structure ([9], [10]) and is implemented by establishing a mapping between the internal data model and BAO classes and properties. The RDF/XML, N3, JSON-LD are different serialization formats of RDF, an example is provided in Annex 6.2. The Excel format is provided for convenience and quick generation of human readable reports of the database content.

The export into ISA (Investigation-Study-Assay) is under development, as described in the eNanoMapper report D3.2. Once completed, the eNanoMapper API will be extended to support the ISA content type and links will be added to the web pages, enabling export of substances and experimental data as well as search results into ISA format, following the (yet unreleased) ISA-JSON specification[2].

[2] https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas

4. CONCLUSIONS

The eNanoMapper prototype database, available at http://data.enanomapper.net/ supports several query types and allows retrieving information of nanomaterials through a REST web services API (http://enanomapper.github.io/API/) and a web browser interface. The demonstration data provided by partners illustrates the capability of the API and the implementation to handle diverse information. It has been used for NanoQSAR modelling through the Jaqpot Quattro Modelling API. The API with JSON serialisation is the current state of the art in web system development and data integration and enables building graphical summaries of the data, JavaScript widgets, custom user interfaces and programmatic interaction. This deliverable report describes a user application providing an intuitive free text search interface, powered by a free text search engine and the eNanoMapper ontology. We also describe how to use the API to send queries through R and JavaScript and build visual summaries. The data download is supported via REST API and allows RDF, CSV, JSON and XLXS download. A blog post, describing how to use the different database search functionalities and an example of a measurement representations in a semantic format are provided as Annexes. The future developments include improvements of the search applications and visualizations based on the received feedback and an automatic generation of ISA archives, compliant with the upcoming ISA-JSON specification.

5. BIBLIOGRAPHY

[1] N. Jeliazkova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliazkov, N. Kochev, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, P. Sopasakis, G. Tsiliki, D. Vorgrimmler, and E. Willighagen, "The eNanoMapper database for nanomaterial safety information," *Beilstein J. Nanotechnol.*, vol. 6, pp. 1609–1634, Jul. 2015.

[2] "Apache SOLR." [Online]. Available: http://lucene.apache.org/solr/. [Accessed: 14-Mar-2016].

[3] J. Hastings, N. Jeliazkova, G. Owen, G. Tsiliki, C. R. Munteanu, C. Steinbeck, and E. Willighagen, "eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment," *J. Biomed. Semantics*, vol. 6, no. 1, 2015.

[4] "The R Project for Statistical Computing." [Online]. Available: https://www.r-project.org/. [Accessed: 14-Mar-2016].

[5] N. Jeliazkova, P. Doganis, B. Fadeel, R. Grafstrom, J. Hastings, V. Jeliazkov, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, G. Tsiliki, D. Vorgrimmler, and E. Willighagen, "The first eNanoMapper prototype: A substance database to support safe-by-design," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 1–9.

[6] S. Gaheen, G. W. Hinkal, S. A. Morris, M. Lijowski, M. Heiskanen, and J. D. Klemm, "caNanoLab: data sharing to expedite the use of nanotechnology in biomedicine," *Comput. Sci. Discov.*, vol. 6, no. 1, p. 014010, Nov. 2013.

[7] S. A. Morris, S. Gaheen, M. Lijowski, M. Heiskanen, and J. Klemm, "Experiences in supporting the structured collection of cancer nanotechnology data using caNanoLab," *Beilstein J. Nanotechnol.*, vol. 6, pp. 1580–1593, Jul. 2015.

[8] N. Jeliazkova, P. Doganis, B. Fadeel, R. Grafstrom, J. Hastings, V. Jeliazkov, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, G. Tsiliki, D. Vorgrimmler, and E. Willighagen, "The first eNanoMapper prototype: A substance database to support safe-by-design," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 1–9.

[9] U. Visser, S. Abeyruwan, U. Vempati, R. P. Smith, V. Lemmon, and S. C. Schürer, "BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results.," *BMC Bioinformatics*, vol. 12, no. 1, p. 257, Jan. 2011.

[10] S. Abeyruwan, U. D. Vempati, H. Küçük-McGinty, U. Visser, A. Koleti, A. Mir, K. Sakurai, C. Chung, J. A. Bittker, P. A. Clemons, S. Brudz, A. Siripala, A. J. Morales, M. Romacker, D. Twomey, S. Bureeva, V. Lemmon, and S. C. Schürer, "Evolving BioAssay Ontology (BAO): modularization, integration and applications," *J. Biomed. Semantics*, vol. 5, no. Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G, p. S5, Jan. 2014.

6. ANNEXES

ANNEX 6.1. HOW TO USE THE ENM DATABASE SEARCH FUNCTION

Friederike Ehrhart, CC-BY 4.0

This blog post shows a few examples of how to use the eNanoMapper database search function to find specific data:

1. how to find data about TiO2 nanoparticles?
2. how to find data of carcinogenicity experiments with nanoparticles?
3. how can we merge data from different assays?
4. which nanomaterials in the database contain PEG?
5. what happens if there is nothing in the database about my material?

6.1.1. HOW TO FIND DATA ABOUT TiO_2 NANOPARTICLES?

Specific nanoparticles can be found by using either nanoparticle name, identifier, composition, or free text search. Entering e.g. TiO2 in the “name search” will give the list in figure 1 of available data. The grey column on the right side allows filtering for different terms like data sources, nanomaterial types and others. Details for the single studies can be found by clicking on the links and the derived results can be downloaded in the form of JSON, CSV, XLSX or RDF (RDF-XML, N3 or JSON-LD) files for further application.

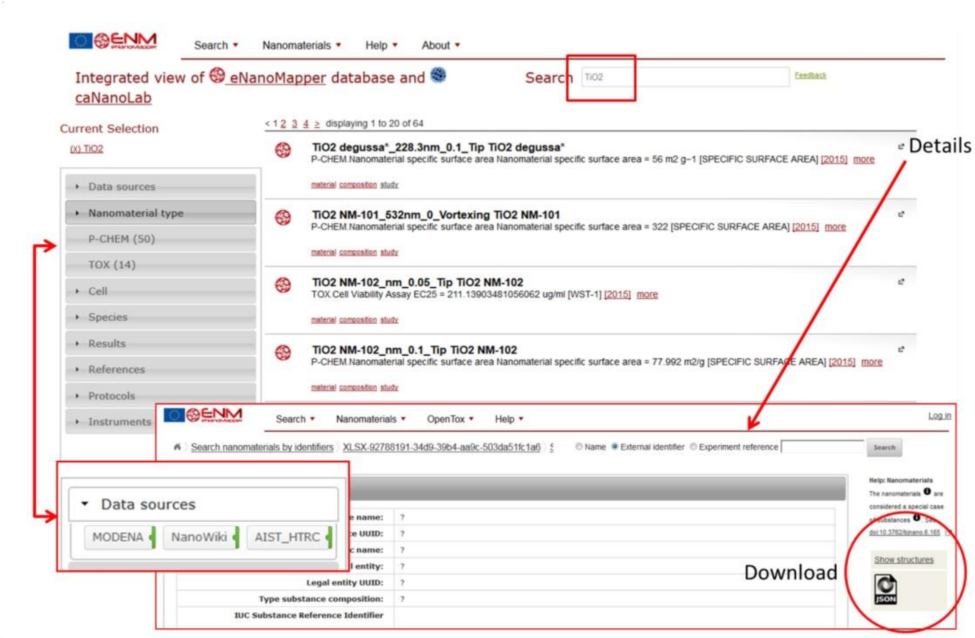


Figure 1: eNM datasets concerning TiO2 nanoparticles.

Figure 6.1.1: eNM datasets concerning TiO2 nanoparticles.

6.1.2. HOW TO FIND DATA OF CARCINOGENICITY EXPERIMENTS WITH NANOPARTICLES?

The eNM database offers the possibility to search specific data containing phys-chem or biological assays (e.g. Search - Search nanomaterials by physchem parameters or biological effects) figure 2. Then, select the assays in question on the checklist at the left side (e.g. Carcinogenicity assay) and apply “update results”. The resulting list will show all database entries with that assay and provide a possibility for download (figure 3).

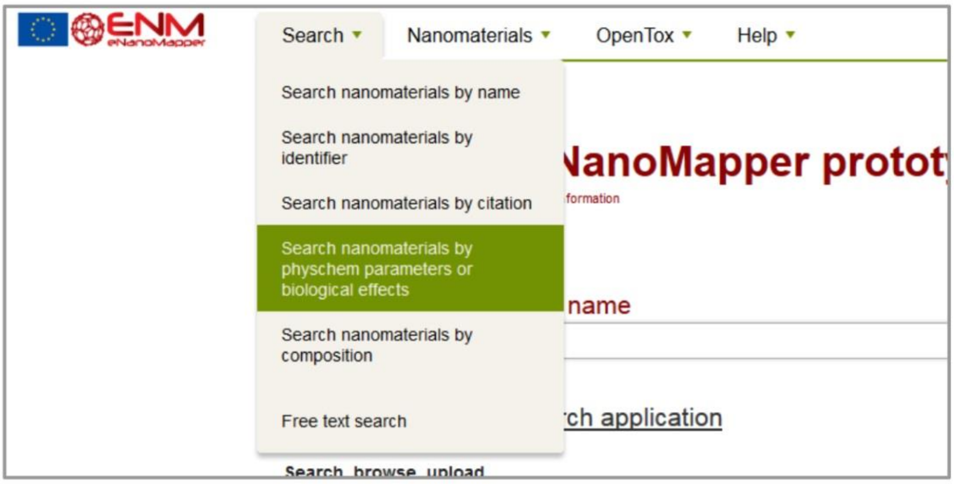


Figure 6.1.2: Search function

6.1.3. HOW CAN WE MERGE DATA FROM DIFFERENT ASSAYS?

Using ontology terms allows merging several assays (including similar experiments). The physchem/biological effect search (see figure 2) allows to filter the results for general terms like cell viability or cytotoxicity which include a variety of different assays. The list is dynamic and the number and the specificity of the assays in the list depends on how the imported data was annotated. For example, if asked for datasets with cell viability endpoints, the current database returns 80 datasets. Searching specifically for LDH assay (lactate dehydrogenase release assay, an indicator for membrane integrity and therefore cell viability), one finds there are 17 datasets from FP7 MARINA and MODENA projects. With more (properly annotated) data being imported, the list will extend. The filter functions in the hit list are currently limited to the textual information available in the table (i.e. substance names and study owners, external identifiers). Desirable would also be terms which allow filtering and ranking for quality measures like data size, data completeness, and date of data production (or publication).

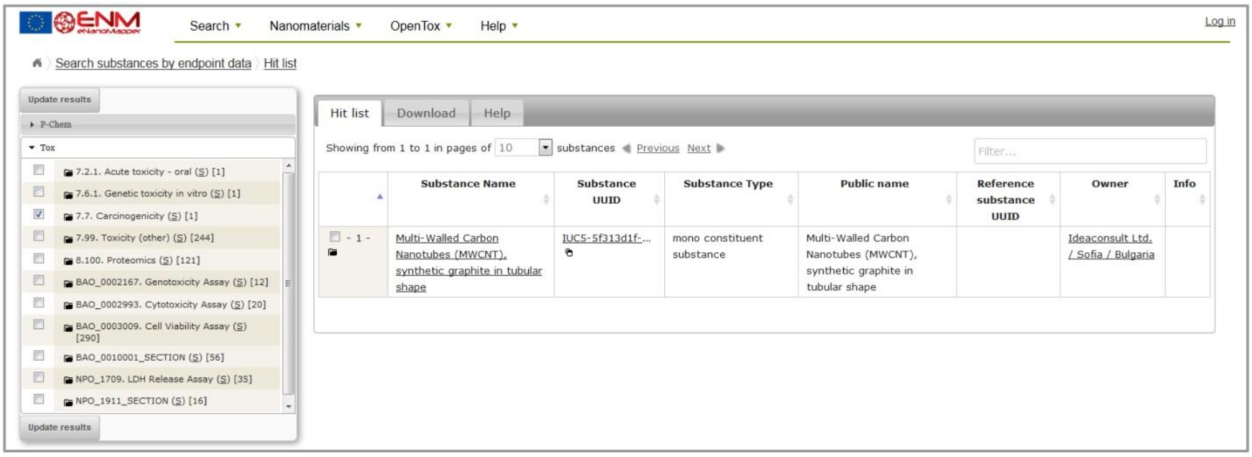


Figure 6.1.3: Result example in the eNM database for carcinogenicity data.

6.1.4. WHICH NANOMATERIALS IN THE DATABASE CONTAIN PEG

To get data of nanomaterials containing a certain material in core, shell or linker material, the search function “nanomaterials by composition” can be used. The hits are labelled (e.g. core or shell). The example in figure 4 shows the currently available data about PEG (polyethylene glycol) containing nanoparticles in the eNM database.

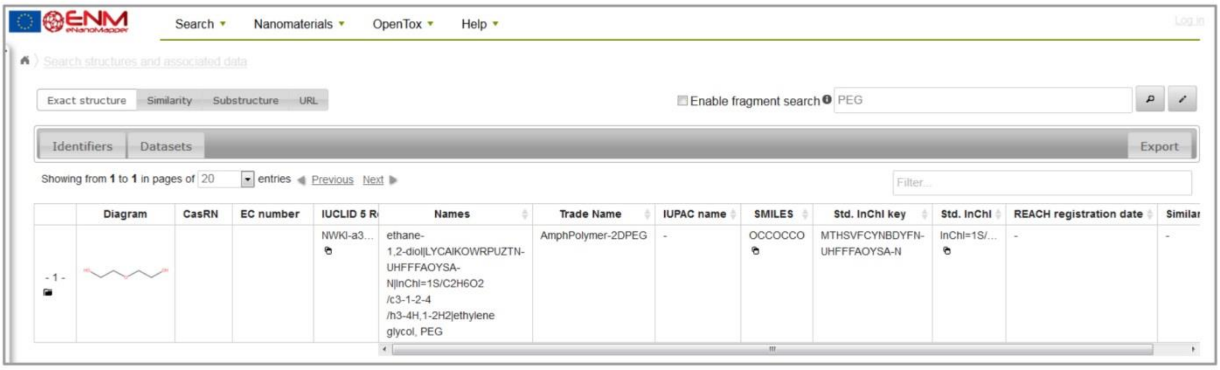


Figure 4: Search by nanomaterials by composition using PEG as search term.

6.1.5. WHAT HAPPENS IF THERE IS NOTHING IN THE DATABASE ABOUT MY MATERIAL?

Looking for material which is not yet in the database (e.g. “kitty litter”) will return “no substances available” as shown in figure 5. Please feel free to upload your data (see this eNanoMapper tutorial on the eNanoMapper website: 3. Data preparation and upload) or contact the eNanoMapper consortium.

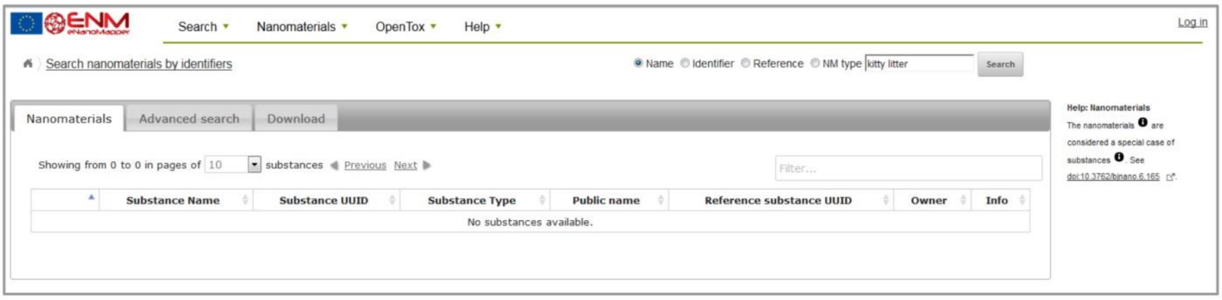
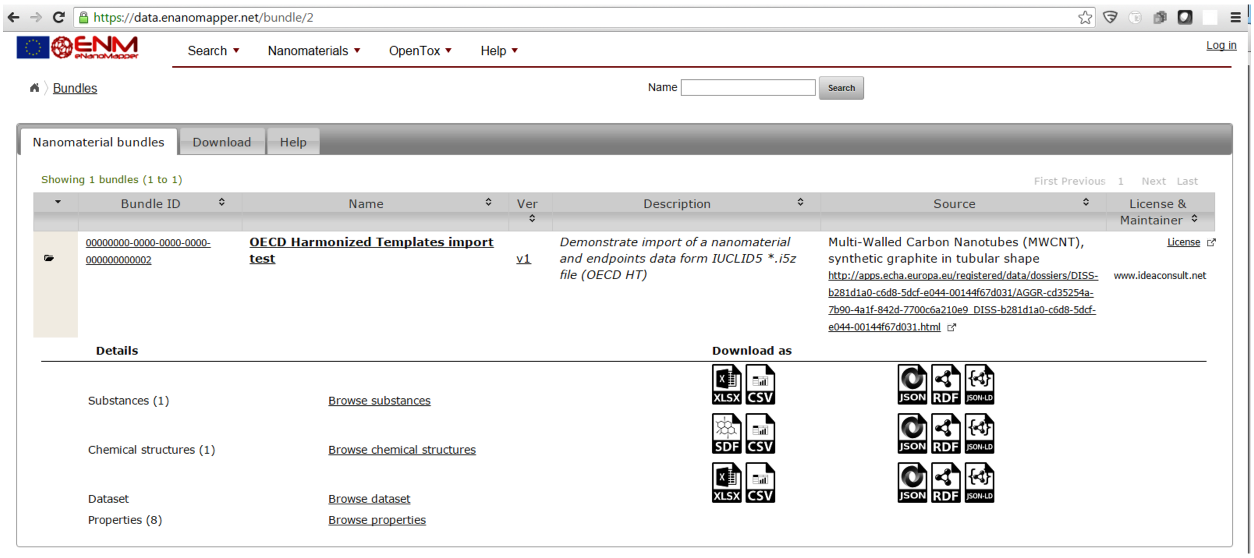


Figure 5: No material available in the database.

ANNEX 6.2. RDF SERIALISATION EXAMPLES

This annex provides example of the RDF/XML serialization of one of the bundles available at http://data.enanomapper.net, namely the OECD Harmonized Templates import test, containing physicochemical and toxicity data on Multi-Walled Carbon Nanotubes. The source of the data was described in D3.1 and [1].



The RDF/XML was retrieved using the following URL
https://data.enanomapper.net/bundle/2/substance?media=application%2Frdf%2Bxml

The retrieved content is listed below.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY this 'https://data.enanomapper.net/'>
  <!ENTITY ot 'http://www.opentox.org/api/1.1#'>
```

```
  <!ENTITY dc 'http://purl.org/dc/elements/1.1/'>
  <!ENTITY enm 'http://purl.enanomapper.org/onto/'>
  <!ENTITY ota 'http://www.opentox.org/algorithmTypes.owl#'>
  <!ENTITY as 'https://data.enanomapper.net/substance/'>
  <!ENTITY otee 'http://www.opentox.org/echaEndpoints.owl#'>
  <!ENTITY bx 'http://purl.org/net/nknouf/ns/bibtex#'>
  <!ENTITY npo 'http://purl.bioontology.org/ontology/npo/'>
  <!ENTITY dcterms 'http://purl.org/dc/terms/'>
  <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
  <!ENTITY obo 'http://purl.obolibrary.org/obo/'>
  <!ENTITY am 'https://data.enanomapper.net/model/'>
  <!ENTITY sio 'http://semanticscience.org/resource/'>
  <!ENTITY ac 'https://data.enanomapper.net/compound/'>
  <!ENTITY owl 'http://www.w3.org/2002/07/owl#'>
  <!ENTITY xsd 'http://www.w3.org/2001/XMLSchema#'>
  <!ENTITY ad 'https://data.enanomapper.net/dataset/'>
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY ag 'https://data.enanomapper.net/algorithm/'>
  <!ENTITY af 'https://data.enanomapper.net/feature/'>
  <!ENTITY bao 'http://www.bioassayontology.org/bao#'>]>
<rdf:RDF
    xmlns:ag="&ag;"
    xmlns:af="&af;"
    xmlns:obo="&obo;"
    xmlns:ot="&ot;"
    xmlns:bx="&bx;"
    xmlns:am="&am;"
    xmlns:otee="&otee;"
    xmlns:dc="&dc;"
    xmlns:bao="&bao;"
    xmlns="&this;"
    xmlns:rdf="&rdf;"
    xmlns:npo="&npo;"
    xmlns:ac="&ac;"
    xmlns:enm="&enm;"
    xmlns:sio="&sio;"
    xmlns:as="&as;"
    xmlns:owl="&owl;"
    xmlns:dcterms="&dcterms;"
    xmlns:xsd="&xsd;"
    xmlns:ota="&ota;"
    xmlns:rdfs="&rdfs;"
    xmlns:ad="&ad;"
  xml:base="&this;">
  <npo:NPO_1235 rdf:about="assay/IUC5-4d6e67a8-4301-4e6a-bafc-20e20fbaf2ad">
    <obo:BAO_0000209>
      <rdf:Description rdf:about="measuregroup/IUC5-4d6e67a8-4301-4e6a-bafc-
20e20fbaf2ad">
        <obo:OBI_0000299>
          <rdf:Description rdf:about="endpoint/ID9">
            <sio:has-unit>m²/g</sio:has-unit>
            <sio:has-value rdf:datatype="&xsd;double">253.0</sio:has-value>
            <rdfs:label>SPECIFIC_SURFACE_AREA</rdfs:label>
            <obo:IAO_0000136>
```

```
              <rdf:Description rdf:about="substance/IUC5-5f313d1f-4129-499c-abbe-
ac18642e2471">
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-f6eb018b-2dc5-462b-
be29-a6ffce0fd4d4">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID0">
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                  </rdf:Description>
                </obo:BFO_0000056>
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-9db5f020-e0fa-47c8-
8e61-3eeb8038e279">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID13">
                        <rdfs:label>solid</rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="interpretation/IUC5-9db5f020-e0fa-
47c8-8e61-3eeb8038e279">
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                        <sio:has-value>inorganic</sio:has-value>
                      </rdf:Description>
                    </obo:OBI_0000299>
                  </rdf:Description>
                </obo:BFO_0000056>
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-74e91cde-2ffa-43ec-
8fcd-5d6407f1602c">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID14">
                        <rdfs:label>DUSTINESS INDEX</rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                  </rdf:Description>
                </obo:BFO_0000056>
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-708684e4-4e3b-49cd-
92ec-9fc2f3a4dc14">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID4">
                        <sio:has-unit>mV</sio:has-unit>
                        <sio:has-value rdf:datatype="&xsd;double"
                        >34.5</sio:has-value>
                        <rdfs:label>ZETA POTENTIAL</rdfs:label>
```

```
                <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
              </rdf:Description>
            </obo:OBI_0000299>
            <obo:OBI_0000299>
              <rdf:Description rdf:about="endpoint/ID3">
                <sio:has-unit>mV</sio:has-unit>
                <sio:has-value rdf:datatype="&xsd;double"
                >21.4</sio:has-value>
                <rdfs:label>ZETA POTENTIAL</rdfs:label>
                <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
              </rdf:Description>
            </obo:OBI_0000299>
            <obo:OBI_0000299>
              <rdf:Description rdf:about="endpoint/ID2">
                <sio:has-unit>mV</sio:has-unit>
                <sio:has-value rdf:datatype="&xsd;double"
                >39.3</sio:has-value>
                <rdfs:label>ZETA POTENTIAL</rdfs:label>
                <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
              </rdf:Description>
            </obo:OBI_0000299>
            <obo:OBI_0000299>
              <rdf:Description rdf:about="endpoint/ID1">
                <sio:has-unit>mV</sio:has-unit>
                <sio:has-value rdf:datatype="&xsd;double"
                >41.4</sio:has-value>
                <rdfs:label>ZETA POTENTIAL</rdfs:label>
                <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
              </rdf:Description>
            </obo:OBI_0000299>
          </rdf:Description>
        </obo:BFO_0000056>
        <obo:BFO_0000056>
          <rdf:Description rdf:about="measuregroup/IUC5-6bdaad41-66c2-4d24-
99ee-8d579243ddd9">
            <obo:OBI_0000299>
              <rdf:Description rdf:about="endpoint/ID6">
                <sio:has-value rdf:datatype="&xsd;string"
                >negative</sio:has-value>
                <rdfs:label>GENOTOXICITY</rdfs:label>
                <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
              </rdf:Description>
            </obo:OBI_0000299>
            <obo:OBI_0000299>
              <rdf:Description rdf:about="endpoint/ID5">
                <sio:has-value rdf:datatype="&xsd;string"
                >negative</sio:has-value>
                <rdfs:label>GENOTOXICITY</rdfs:label>
```

```
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                  </rdf:Description>
                </obo:BFO_0000056>
                <obo:BFO_0000056 rdf:resource="measuregroup/IUC5-4d6e67a8-4301-4e6a-
bafc-20e20fbaf2ad"/>
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-4bca14fc-3e43-435e-
a06e-d924fc77bded">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID12">
                        <sio:has-unit>nm</sio:has-unit>
                        <sio:has-value rdf:datatype="&xsd;double"
                        >12.7</sio:has-value>
                        <rdfs:label>PARTICLE SIZE.D90</rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID11">
                        <rdfs:label>MASS MEDIAN AERODYNAMIC DIAMETER</rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID10">
                        <sio:has-unit>nm</sio:has-unit>
                        <sio:has-value rdf:datatype="&xsd;double"
                        >3.0</sio:has-value>
                        <rdfs:label>MASS MEDIAN DIAMETER</rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                  </rdf:Description>
                </obo:BFO_0000056>
                <obo:BFO_0000056>
                  <rdf:Description rdf:about="measuregroup/IUC5-49179f2f-867c-4e7c-
8fd1-d09eb23c4c3d">
                    <obo:OBI_0000299>
                      <rdf:Description rdf:about="endpoint/ID8">
                        <sio:has-unit>mg/kg bw</sio:has-unit>
                        <sio:has-value rdf:datatype="&xsd;double"
                        >5000.0</sio:has-value>
                        <rdfs:label>LD50 cut-off </rdfs:label>
                        <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                      </rdf:Description>
                    </obo:OBI_0000299>
                    <obo:OBI_0000299>
```

```
                        <rdf:Description rdf:about="endpoint/ID7">
                          <sio:has-unit>mg/kg bw</sio:has-unit>
                          <sio:has-value rdf:datatype="&xsd;double"
                          >2000.0</sio:has-value>
                          <rdfs:label>LDLo</rdfs:label>
                          <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                        </rdf:Description>
                      </obo:OBI_0000299>
                      <obo:OBI_0000299>
                        <rdf:Description rdf:about="interpretation/IUC5-49179f2f-867c-
4e7c-8fd1-d09eb23c4c3d">
                          <obo:IAO_0000136 rdf:resource="substance/IUC5-5f313d1f-4129-
499c-abbe-ac18642e2471"/>
                          <sio:has-value>practically nontoxic</sio:has-value>
                        </rdf:Description>
                      </obo:OBI_0000299>
                    </rdf:Description>
                  </obo:BFO_0000056>
                  <dcterms:source>
                    <rdf:Description rdf:about="owner/IUC5-354430c7-746e-450e-9722-
09788eb90a29">
                      <dcterms:title>Ideaconsult Ltd. / Sofia /
Bulgaria</dcterms:title>
                    </rdf:Description>
                  </dcterms:source>
                </rdf:Description>
              </obo:IAO_0000136>
            </rdf:Description>
          </obo:OBI_0000299>
        </rdf:Description>
      </obo:BAO_0000209>
      <obo:BAO_0002846>
        <rdf:Description rdf:about="protocol/3231603C">
          <dc:title>DIN66131</dc:title>
        </rdf:Description>
      </obo:BAO_0002846>
      <dc:title>Specific surface area.001</dc:title>
    </npo:NPO_1235>
    <bao:BAO_0002167 rdf:about="assay/IUC5-6bdaad41-66c2-4d24-99ee-8d579243ddd9">
      <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-6bdaad41-66c2-4d24-99ee-
8d579243ddd9"/>
      <dcterms:source>
        <rdf:Description rdf:about="reference/3706B8E3">
          <dc:title>Thurnherr, T. et al.</dc:title>
        </rdf:Description>
      </dcterms:source>
      <obo:BAO_0002846 rdf:resource="protocol/0F26FAB2"/>
      <dc:title>Exp Supporting Genetic toxicity in vitro.008</dc:title>
    </bao:BAO_0002167>
    <obo:CHMO_0002119 rdf:about="assay/IUC5-4bca14fc-3e43-435e-a06e-d924fc77bded">
      <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-4bca14fc-3e43-435e-a06e-
d924fc77bded"/>
      <obo:BAO_0002846>
```

```
      <rdf:Description rdf:about="protocol/A9046FFE">
        <dc:title>Transmission Electron Microscopy (TEM)</dc:title>
      </rdf:Description>
    </obo:BAO_0002846>
    <dc:title>Particle size distribution (Granulometry).001</dc:title>
  </obo:CHMO_0002119>
  <enm:ENM_9000003 rdf:about="assay/IUC5-74e91cde-2ffa-43ec-8fcd-5d6407f1602c">
    <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-74e91cde-2ffa-43ec-8fcd-
5d6407f1602c"/>
    <obo:BAO_0002846>
      <rdf:Description rdf:about="protocol/DCF1BC7D">
        <dc:title>DIN 33897-2 (2002) Workplace atmospheres - Determination of the
dustiness of bulk materials - Part 2:Continuous Drop (note: SUPERSEDED by EN
1505:2006)</dc:title>
      </rdf:Description>
    </obo:BAO_0002846>
    <dc:title>Dustiness.001</dc:title>
  </enm:ENM_9000003>
  <enm:ENM_0000029 rdf:about="assay/IUC5-f6eb018b-2dc5-462b-be29-a6ffce0fd4d4">
    <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-f6eb018b-2dc5-462b-be29-
a6ffce0fd4d4"/>
    <dcterms:source>
      <rdf:Description rdf:about="reference/AEA4F543">
        <dc:title>Muller, J. et al.</dc:title>
      </rdf:Description>
    </dcterms:source>
    <obo:BAO_0002846 rdf:resource="protocol/7F6AED3C"/>
    <dc:title>Carcinogenicity.001</dc:title>
  </enm:ENM_0000029>
  <obo:BAO_0000015 rdf:about="assay/IUC5-9db5f020-e0fa-47c8-8e61-3eeb8038e279">
    <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-9db5f020-e0fa-47c8-8e61-
3eeb8038e279"/>
    <dcterms:source>
      <rdf:Description rdf:about="reference/4D5BA95A">
        <dc:title>Ragot J, Voetz M </dc:title>
      </rdf:Description>
    </dcterms:source>
    <obo:BAO_0002846>
      <rdf:Description rdf:about="protocol/88CD6B80">
        <dc:title></dc:title>
      </rdf:Description>
    </obo:BAO_0002846>
    <dc:title>Appearance/physical state/colour.001</dc:title>
  </obo:BAO_0000015>
  <npo:NPO_1302 rdf:about="assay/IUC5-708684e4-4e3b-49cd-92ec-9fc2f3a4dc14">
    <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-708684e4-4e3b-49cd-92ec-
9fc2f3a4dc14"/>
    <obo:BAO_0002846>
      <rdf:Description rdf:about="protocol/2EE61FE6">
        <dc:title></dc:title>
      </rdf:Description>
    </obo:BAO_0002846>
    <dc:title>Zeta potential.001</dc:title>
  </npo:NPO_1302>
```

```
  <enm:ENM_0000020 rdf:about="assay/IUC5-49179f2f-867c-4e7c-8fd1-d09eb23c4c3d">
    <obo:BAO_0000209 rdf:resource="measuregroup/IUC5-49179f2f-867c-4e7c-8fd1-
d09eb23c4c3d"/>
    <obo:BAO_0002846>
      <rdf:Description rdf:about="protocol/EF511EF5">
        <dc:title>OECD Guideline 423 (Acute Oral toxicity - Acute Toxic Class
Method)</dc:title>
      </rdf:Description>
    </obo:BAO_0002846>
    <dc:title>Acute toxicity: oral.001</dc:title>
  </enm:ENM_0000020>
</rdf:RDF>
```

The RDF consists of a set of triples, representing interlinked set of objects, mostly defined by the BAO ontology. The visualization was generated using the W3C validator and graph generator, available at https://www.w3.org/RDF/Validator/rdfval (see Fig. 6.2.1).

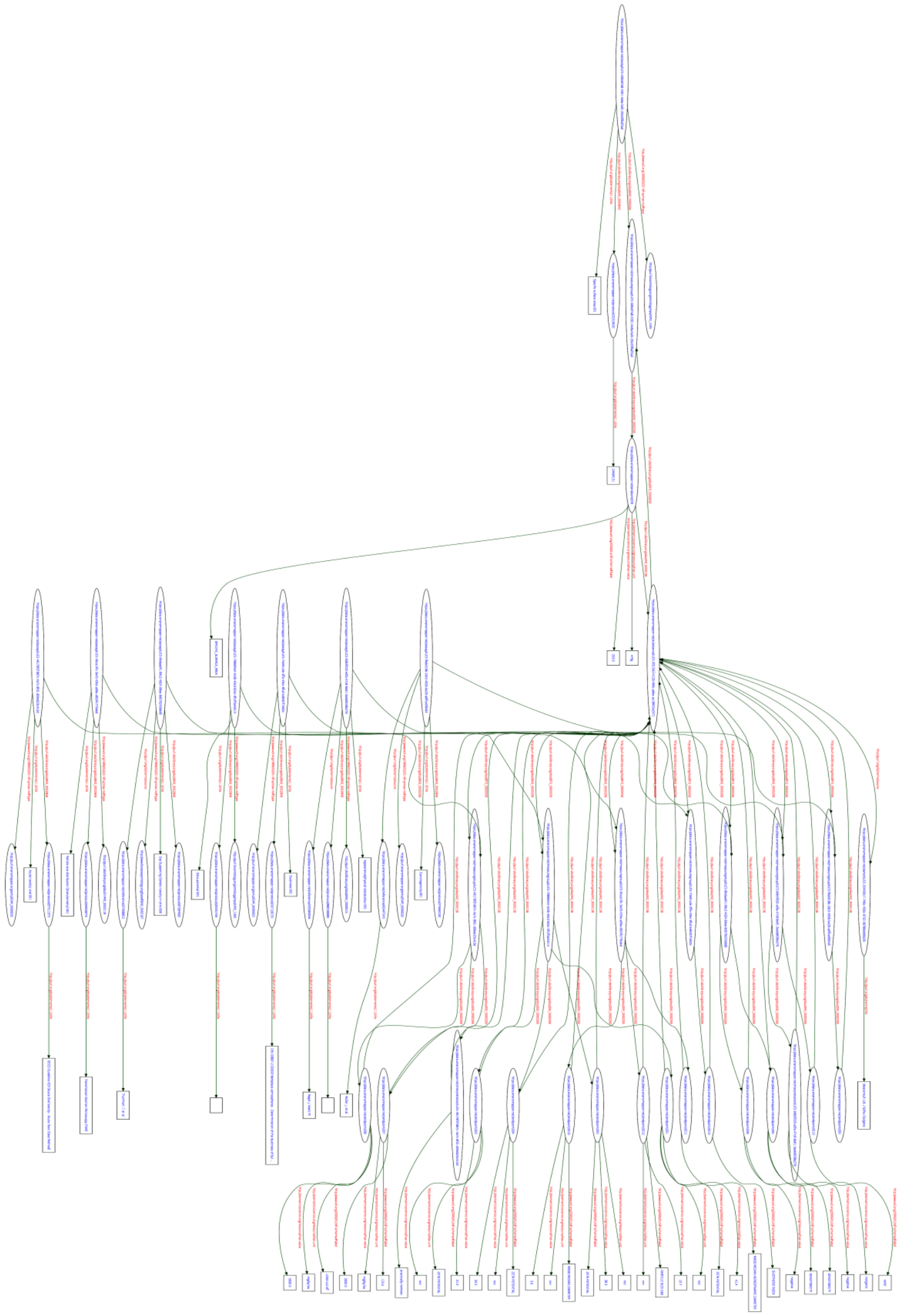


Figure 6.2.1. A linked graph describing physicochemical characterization and toxicity data.