How to store nanomaterial safety data : meet eNanoMapper database

Nikolay T. Kochev[a,c], Roland Grafström [b], Nina Jeliazkova [c*]

a) University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry, 24 Tsar Assen Str., 4000 Plovdiv, Bulgaria
b) Institute of Environmental Medicine, Karolinska Institutet, P.O. Box 210, SE-171 77 Stockholm, Sweden
c) Ideaconsult Ltd, 4 Angel Kanchev Str., 1000 Sofia, Bulgaria
* To whom correspondence should be addressed. e-mail: jeliazkova.nina@gmail.com , twitter: @10705013

http://enanomapper.net

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 604134.

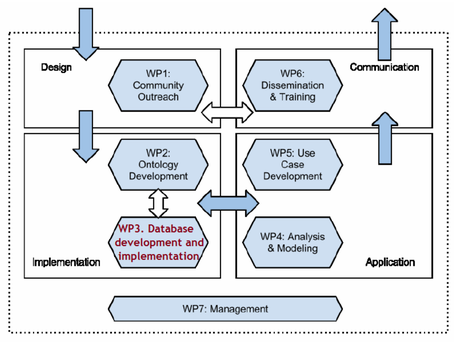
eNanoMapper FP7 project

Objectives:

- Develop an ontology and database unifying information about nanomaterial safety
- Cover the full lifecycle from manufacturing to environmental decay or accumulation
- Pan-European project, 8 partners
- Ontology growth through community and re-use

Vision:

- Based on OpenTox API
- Open Source implementations
- Bridging with data analysis tools
- Wide range of exchange data formats



Challenges

- Diverse requirements, posed by the nanotechnology community;
- Data representation and integration challenges mainly due to data complexity and provenance.

- **Physicochemical identity**

Different analytic techniques, manufacturing conditions, batch effects, mixtures, impurities, size distribution, differences in the amount of surface modification, etc.

- **Biological identity**

Wide variety of measurements, toxicity pathways, effects of ENM coronas, modes-of-action, interactions (cell lines, assays).

- **Data formats, Provenance, Visualisation**

From raw data to study summaries for regulatory purposes; linking with experimental protocols; user friendly visualisation.

- **Support for data analysis**

Requires “spreadsheet” or matrix view of data. The experimental data in the public datasets is usually not in a form appropriate for modelling. Standardisation in these sources is specific to each database. Even in curated collections the preparation of data for modelling is not a straightforward exercise (e.g. the experimental values can be merged in many different ways into a matrix, depending on which experimental protocols and conditions are considered similar; also there could be multiple values due to replicates or similar experiments)

Chemical structures, substances, nanomaterials

Substance definition:

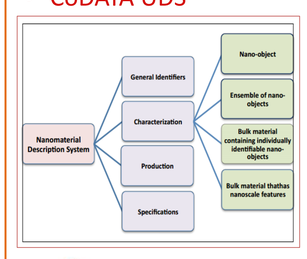
- ***NanoParticle Ontology (NPO):*** a Nanomaterial (NPO_199) is an equivalent class to chemical substance (NPO_1973) one of (nano-object, nanoparticle, engineered nanomaterial, nanostructured material, nanoparticle formulation). The chemical substance itself is a subclass of a chemical entity (NPO_1972).
- **DOI: 10.1007/s11051-013-1455-2 , . J. Nanoparticle Res. 2013, 15 :** Compares the definition of the terms “*substance*” and “*material*” are in ISO, REACH and general science definitions of the terms. The paper notes the OECD HT definition of “reference substances” is very similar to the definition of the term “reference material”.
- **REACH Guidance:** “Chemical substance, a material with a definite chemical composition”. The definition of a substance encompasses all forms of substances and materials on the market, including nanomaterials; and may have complex composition.

Measurements

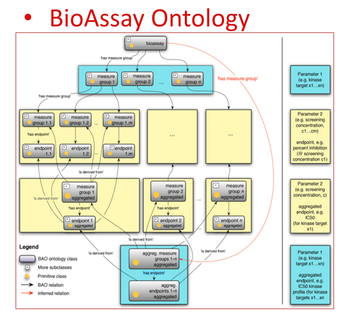
- ***NanoParticle Ontology (NPO):*** distinguishes between endpoint of measurement and assay used to measure the endpoint, where the details of the assay could be specified
- **DOI: 10.1007/s11051-013-1455-2, . J. Nanoparticle Res. 2013, 15 :** “test” and “measurement” terms
- **CoDATA Universal Description System :** requires specification of how particular property is measured
- **ISA-Tab-Nano:** allows defining the qualities measured and detailed protocol conditions and instruments
- **BioAssay Ontology (BAO) :** Experimental data is organized in ``measure groups''. A measure group can be annotated with an endpoint, screened entity (e.g. ENM), assay method and participants. A bioassay may contain multiple measurement groups.
- **OECD Harmonized Templates:** > 100 endpoint specific XML schema

Existing data models

- CoDATA UDS



- BioAssay Ontology



- OECD harmonized Templates
- ISA-Tab/ISA-Tab-Nano

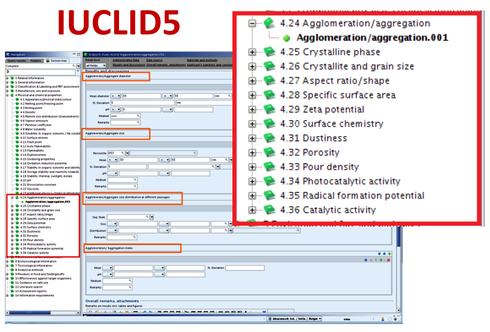
- Nanomaterials
  - Core
  - Coating(s)
  - Linkage
  - Impurities
  - Components, internal structure, etc.
- Typical assay description
  - Property – value (range of values) – units *(Excel templates)*
- More complex description:
  - Experimental graph

- Commonalities:
  - Materials sample
  - Protocols, protocol parameters
  - Experimental conditions
  - Readouts
    - Measurements,
    - Measurement groups,
    - Raw data, derived data

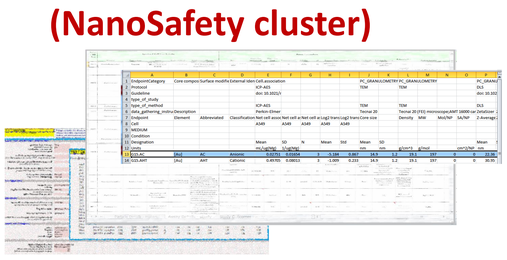
eNanoMapper Data Model

Supported import formats

- **OECD Harmonized Templates, IUCLID5**



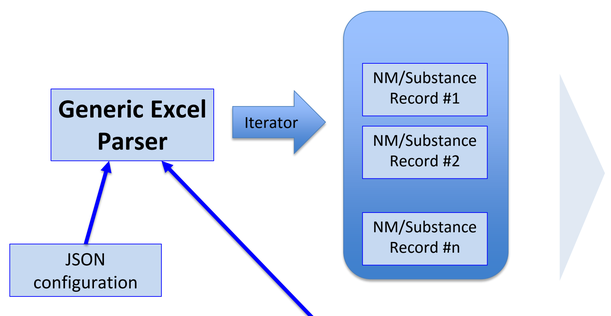
- **NanoWiki RDF (KI, MU)**
- **Custom spreadsheets (NanoSafety cluster)**

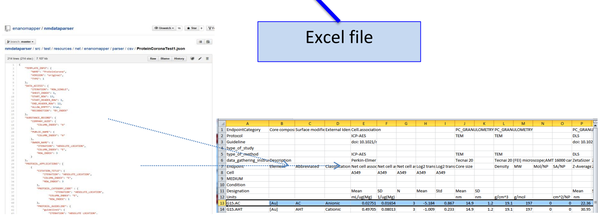


- **ISA-Tab** isatab
- **ISA-Tab-Nano**
- **…**

(under development)

Configurable Excel Parser for custom spreadsheets

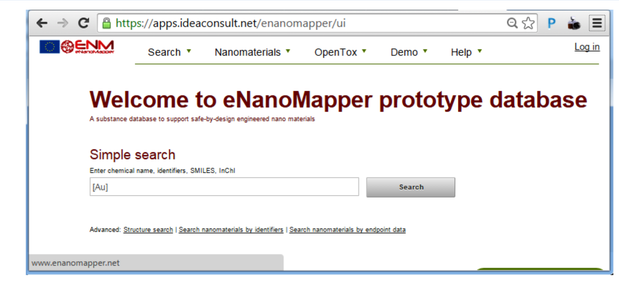




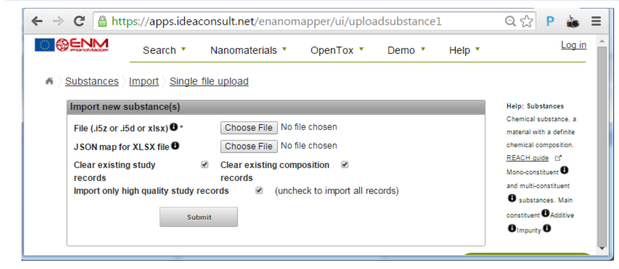
https://github.com/enanomapper/nmdataparser

- Upload of ENM , physchem and assay data from various spreadsheet based sources
- Universal approach (if possible)
- Configuration by an external JSON file
- Code reusability
- Encapsulation/hiding of the original source data details (if needed)

eNanoMapper prototype database http://data.enanomapper.net

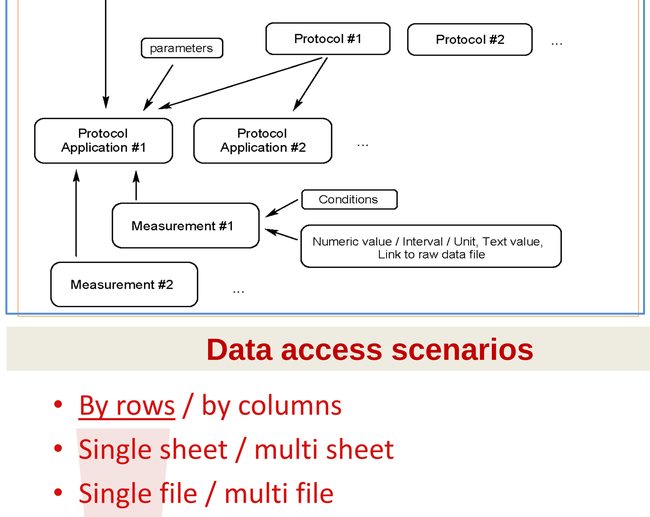


File upload (web form)



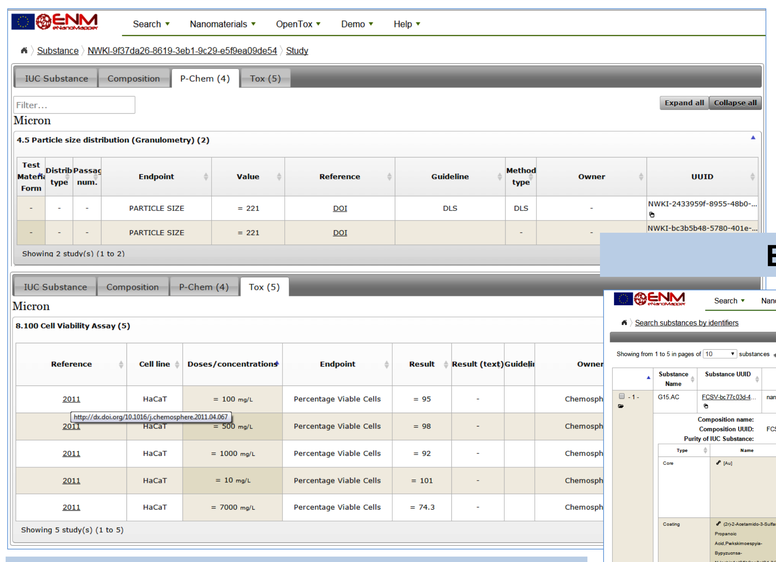
Substance API

http://enanomapper.github.io/API/

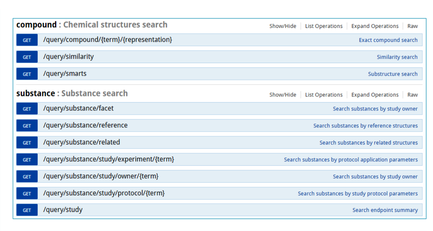


eNanoMapper Data Model

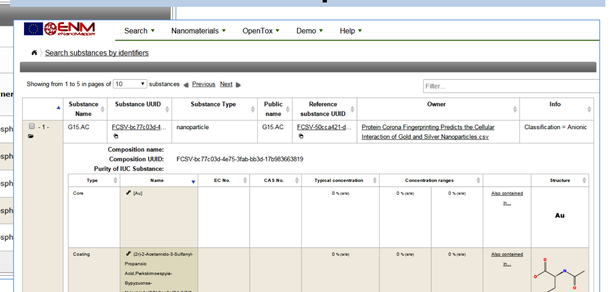
Physicochemical and toxicity data



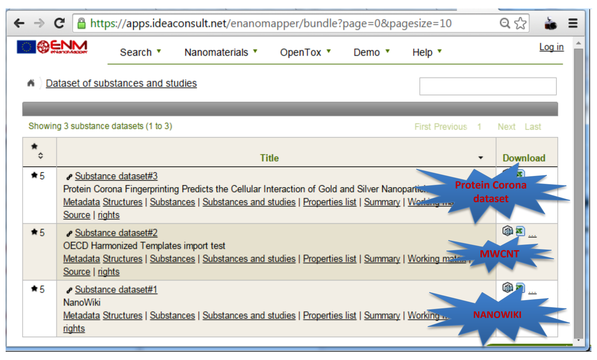
Search API



ENM Components

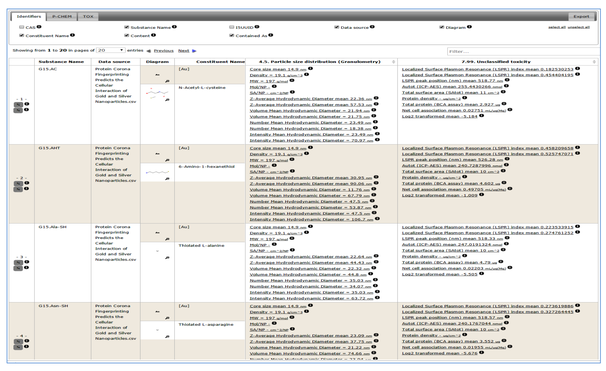


Data collections



- Prepare the ENM characterisation and assay data using your own templates;
- The Excel parser enables converting the input templates into the eNanoMapper data model and upload into the DB;
- Search and explore the eNanoMapper DB;
- Format convertors by exporting the data model into different formats (under dev)
- Ontology annotation (under development)

Matrix view



Data access scenarios

- By rows / by columns
- Single sheet / multi sheet
- Single file / multi file
- Static location definition / Dynamic location definition

- *ROW_SINGLE*

12	Units				mL/ug(Mg)	1/ug(Mg)
13	G15.AC	[Au]	AC	Anionic	0.02751	0.01654	3
14	G15.AHT	[Au]	AHT	Cationic	0.49705	0.08013	3

- *ROW_MULTI_DYNAMIC*

MNT56	4	Published	Griffitt	Robert	Department of Coastal Sciences, University of South Missis
			Hyndman	Kelly	Department of Zoology, University of Florida, Gainesville Fl
			Denslow	Nancy	Department of Coastal Sciences, University of South Missis
			Barber	Davis	Department of Zoology, University of Florida, Gainesville Fl
MTN82	6	Published	Zhao	Chun-Mei	Division of Life Science, The Hong Kong University of Scien Kong, China
			Wang	Wen-Xiong	Division of Life Science, The Hong Kong University of Scien Kong, China

- *ABSOLUTE_LOCATION*

Data is always taken from a particular cell

Mapping the Excel layout into the data model by JSON configuration

Each row represents a substance record

```
"DATA_ACCESS": {
    "ITERATION": "ROW_SINGLE",
    "SHEET_INDEX": 1,
    "START_ROW": 13,
    "START_HEADER_ROW": 1,
    "END_HEADER_ROW": 12,
    "ALLOW_EMPTY": true,
    "RECOGNITION": "BY_INDEX"
},
```

The data locations are specified by the excel column indices

	A	B	C	D	E	F	G
1	EndpointCa	Core compc	Surface mo	External Ide	Cell.association		
2	Protocol				ICP-AES		
3	Guideline				doi: 10.1021/nn4		
4	type_of_study						
5	type_of_method				ICP-AES		
6	data_gathe	Description			Perkin-Elmer		
7	Endpoint	Element	Abbreviate	Classificatic	Net cell associatic	Net cell ass	Net cell ass
8	Cell				A549	A549	A549
9	MEDIUM						
10	Condition						
11	Designation				Mean	SD	N
12	Units				mL/ug(Mg)	1/ug(Mg)	
13	G15.AC	[Au]	AC	Anionic	0.02751	0.01654	3
14	G15.AHT	[Au]	AHT	Cationic	0.49705	0.08013	3

```
"SUBSTANCE_RECORD": {
    "COMPANY_UUID": {
        "COLUMN_INDEX": "A"
    },
    "PUBLIC_NAME": {
        "COLUMN_INDEX": "A"
    },
    "OWNER_NAME": {
        "ITERATION": "ABSOLUTE_LOCATION",
        "COLUMN_INDEX": "E",
        "ROW_INDEX": 3
    },
    "SUBSTANCE_TYPE": {
        "ITERATION": "JSON_VALUE",
        "JSON_VALUE": "Nanoparticle"
    },
    "EXTERNAL_IDENTIFIERS": [
        {
            "ID": {
                "COLUMN_INDEX": "D",
                "ITERATION": "ROW_SINGLE"
            },
            "TYPE": {
                "ITERATION": "ABSOLUTE_LOCATION",
                "COLUMN_INDEX": "D",
                "ROW_INDEX" : 7
            }
        }
```


References : DOI:10.1109/BIBM.2014.6999367; DOI:10.1186/s13326-015-0005-5