eNanoMapper data management: current status

"eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment"

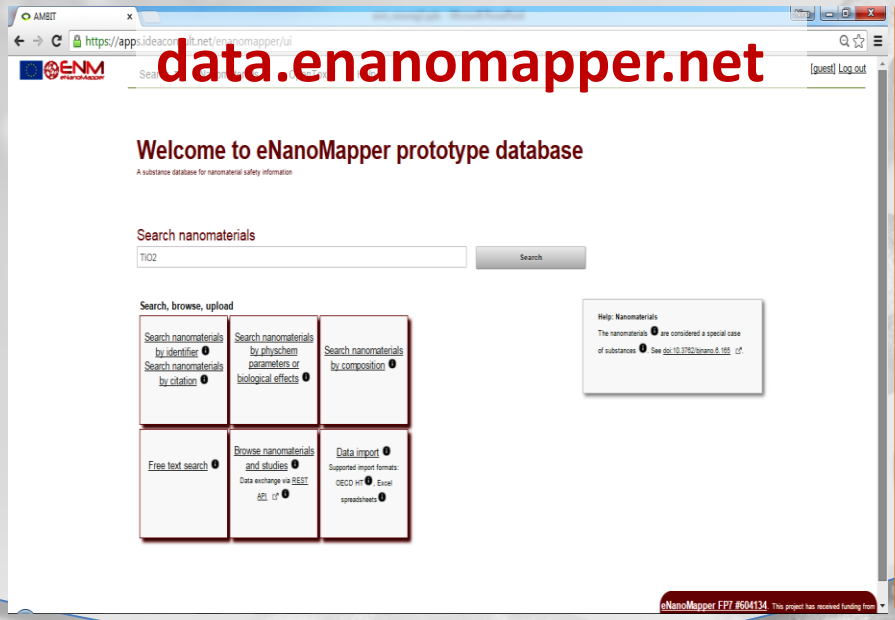
- Grant Agreement: 604134
- Duration: 36 months (1 Feb 2014 – 31 Jan 2017)

WP3. Database development and implementation

NMP.2013.1.3-2 Nanomaterials safety assessment: Ontology, database(s) for modelling and risk assessment

Expected impact:

(i) ... (ii) ... (iii) implementation of the database structure with all of the necessary provisions for data protection, data sharing, data quality assurance, searchability, tailored interfaces for different needs and usages, comparability and cross-talk with other databases;



eNanoMapper Vision:

- Based on OpenTox API
- Web services (multiple)
- Open Source implementations
- Bridging with data analysis tools
- Multiple data exchange format
  - ISA-TAB, semantic formats,
  - OECD HT
  - formats for machine learning packages, etc.

Databases and NM specific challenges

Database

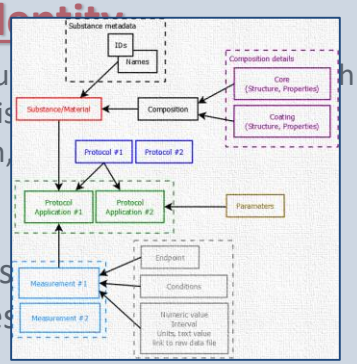
From Wikipedia, the free encyclopedia

A **database** is an organized collection of data.[1] It is the collection of schemes, tables, queries, reports, views and other objects. The data is typically organized to model aspects of reality in a way that supports processes requiring information, such as modelling the availability of rooms in hotels in a way that supports finding a hotel with vacancies.

A **database management system** (**DBMS**) is a computer software application that interacts with the user, other applications, and the database itself to capture and analyze data. A general-purpose DBMS is designed to allow the definition, creation, querying, update, and

- **Physico-chemical identity**

Different analytic techniques, manu… effects, mixtures, impurities, size dis… the amount of surface modification,…



- **Biological identity**

Wide variety of measurements… effects of ENM coronas, modes… interactions (cell lines, assays).

- **Processes requiring information**

From raw data (science) to study summaries for regul… (OECD templates); knowledge, experimental protocols; risk assessment; grouping, safety-by-design

- **Different views of the data**

Requires "spreadsheet" or matrix view of data. The experimental data in the public datasets is usually not in a form appropriate for modelling (merging multiple values, conditions, similar experiments into matrix form is a challenge).

Application programming interface (API)

Different views of the data (e.g. spreadsheet vs "substance dossier")

eNanoMapper

Data sources
1. Excel spreadsheet examples

1) Data transfer between FP7 NANoREG and eNanoMapper (WP3)

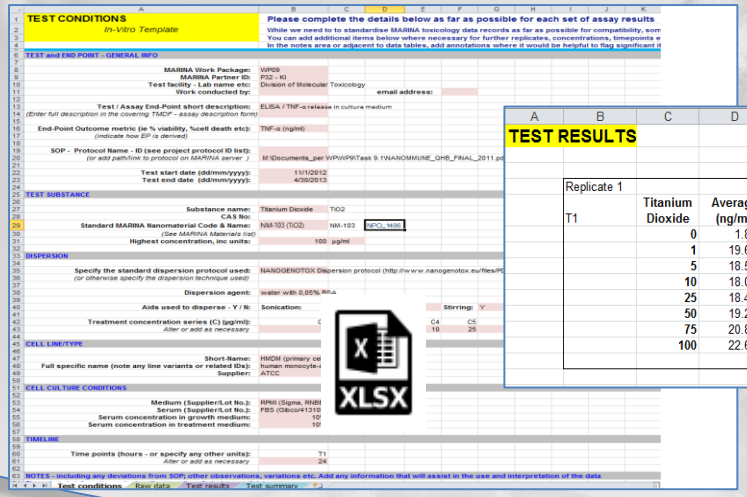
Karolinska Institutet , Prof. Bengt Fadeel lab (NANoREG WP5)

- cytotoxicity assessment of the entire panel of NANoREG nanomaterials
- All 19 nanomaterials were obtained from the JRC nanomaterial repository

2) Data transfer between FP7 MARINA and eNanoMapper (WP3)

Karolinska Institutet , Prof. Bengt Fadeel lab

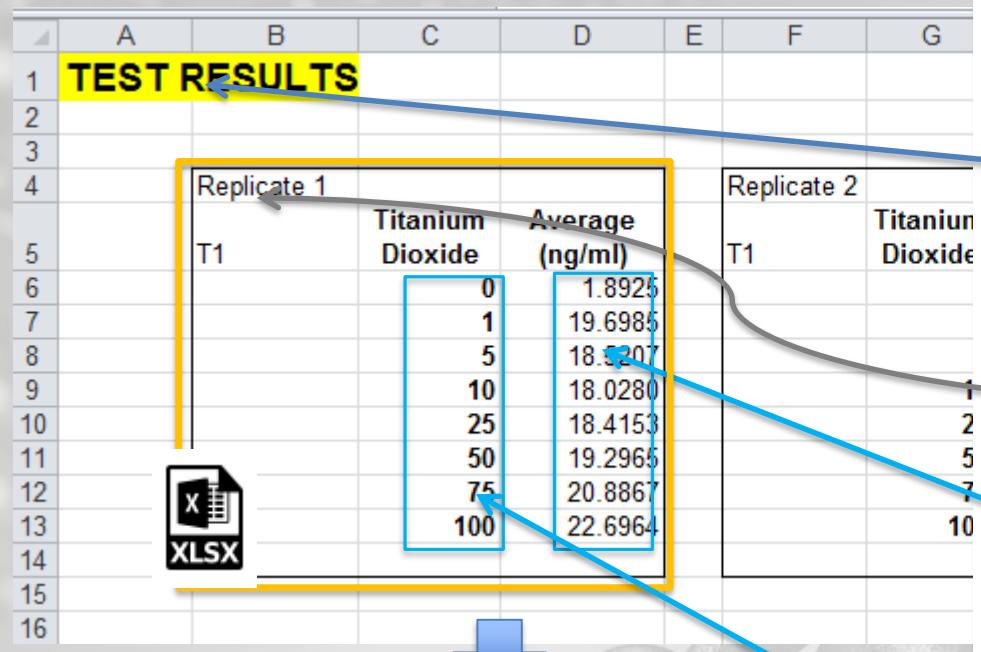
- L. Farcal, F. Torres Andón, L. Di Cristo, B. M. Rotoli, O. Bussolati, E. Bergamaschi, A. Mech, N. B. Hartmann, K. Rasmussen, J. Riego-Sintes, J. Ponti, A. Kinsner-Ovaskainen, F. Rossi, A. Oomen, P. Bos, R. Chen, R. Bai, C. Chen, L. Rocks, N. Fulton, B. Ross, G. Hutchison, L. Tran, S. Mues, R. Ossig, J. Schnekenburger, L. Campagnolo, L. Vecchione, A. Pietroiusti, and B. Fadeel, "Comprehensive In Vitro Toxicity Testing of a Panel of Representative Oxide Nanomaterials: First Steps towards an Intelligent Testing Strategy," *PLoS One*, vol. 10, no. 5, p. e0127174, May 2015.



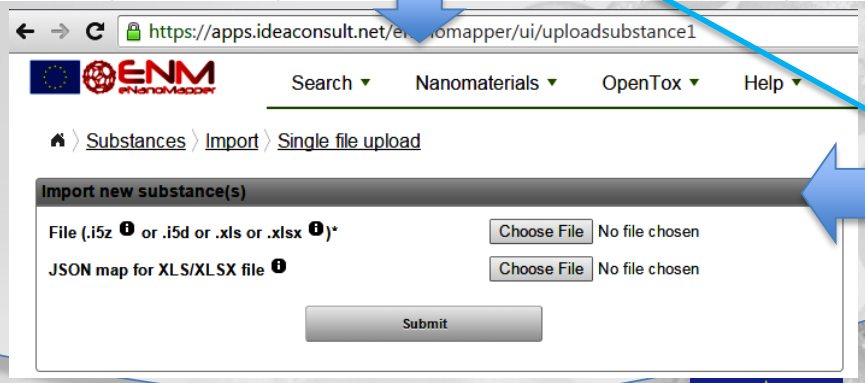
TEST RESULTS

Replicate 1			Replicate 2			Replicate 3			Replicate 4		
T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)
	0	1.8925		0	-0.3425		0	2.189		0	9.502
	1	19.6985		1	5.9887		1	2.474		1	8.808
	5	18.5207		5	5.7696		5	2.002		5	8.957
	10	18.0280		10	5.3011		10	1.786		10	8.306
	25	18.4153		25	5.1602		25	1.570		25	8.244
	50	19.2965		50	5.2292		50	1.735		50	9.507
	75	20.8867		75	6.1564		75	1.290		75	14.416
	100	22.6964		100	7.2811		100	1.206		100	15.438

Mapping the spreadsheet content into the data model



	A	B	C	D	E	F	G
1	TEST RESULTS						
2							
3							
4		Replicate 1				Replicate 2	
5		T1	Titanium Dioxide	Average (ng/ml)		T1	Titanium Dioxide
6			0	1.8925			
7			1	19.6985			
8			5	18.5207			
9			10	18.0280			1
10			25	18.4153			2
11			50	19.2965			5
12			75	20.8867			7
13			100	22.6964			10
14							
15							
16							



Mapping the spreadsheet content into the data model

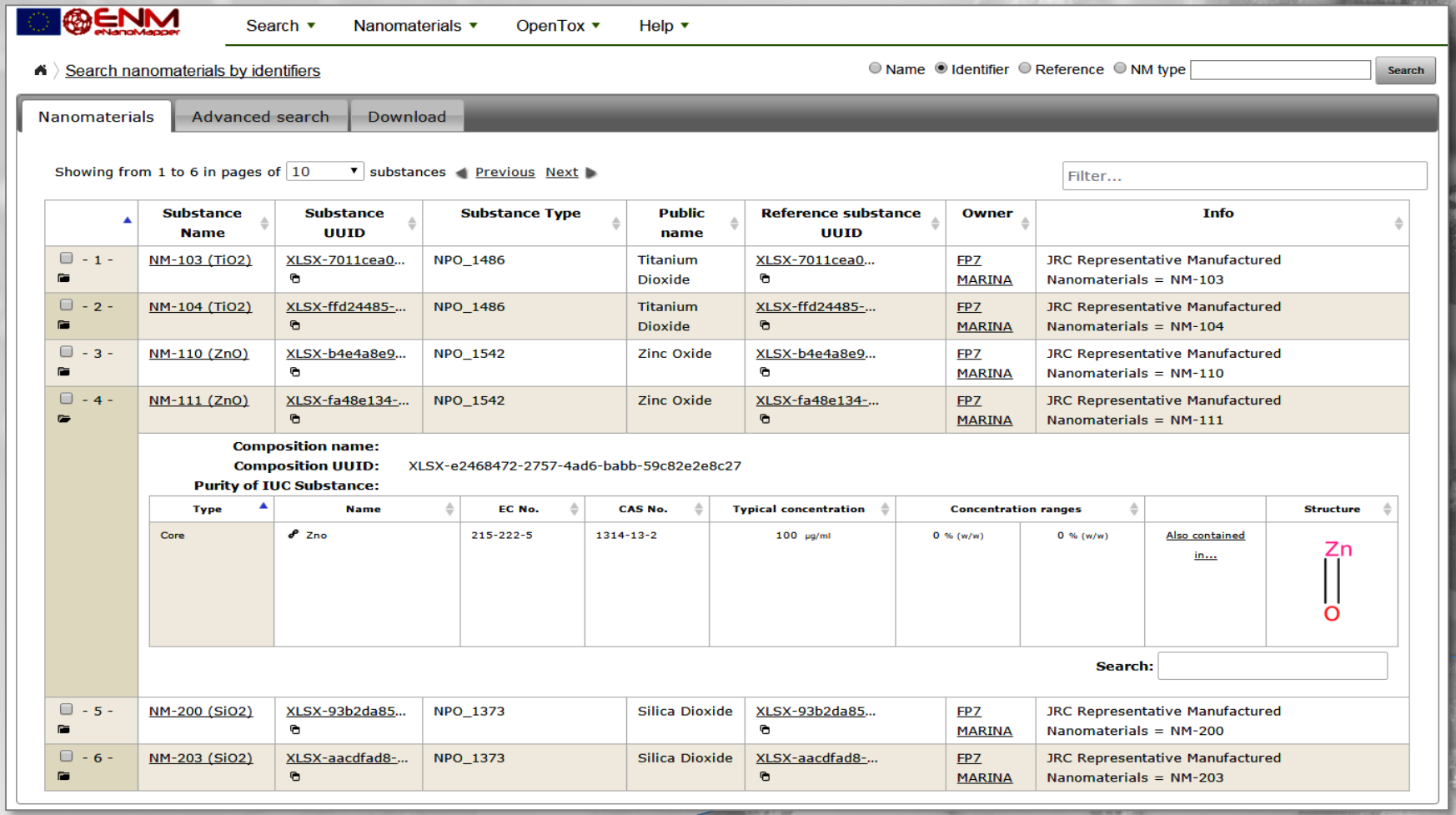
Discussion points

- Assay ontology annotation
- Should we include Pos/ Neg controls?
- Is this study published – what would be the correct reference?
- Is there published protocol to refer to?
- What exactly is measured?
- Cell line
- Any phys-chem data ?

through JSON configuration

eNanoMapper

Data imported (MARINA/KI)



Search ▾ Nanomaterials ▾ OpenTox ▾ Help ▾

Search nanomaterials by identifiers

Name Identifier Reference NM type Search

Nanomaterials | Advanced search | Download

Showing from 1 to 6 in pages of 10 substances Previous Next

Filter...

	Substance Name	Substance UUID	Substance Type	Public name	Reference substance UUID	Owner	Info
- 1 -	NM-103 (TiO2)	XLSX-7011cea0...	NPO_1486	Titanium Dioxide	XLSX-7011cea0...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-103
- 2 -	NM-104 (TiO2)	XLSX-ffd24485-...	NPO_1486	Titanium Dioxide	XLSX-ffd24485-...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-104
- 3 -	NM-110 (ZnO)	XLSX-b4e4a8e9...	NPO_1542	Zinc Oxide	XLSX-b4e4a8e9...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-110
- 4 -	NM-111 (ZnO)	XLSX-fa48e134-...	NPO_1542	Zinc Oxide	XLSX-fa48e134-...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-111

Composition name:
Composition UUID: XLSX-e2468472-2757-4ad6-babb-59c82e2e8c27
Purity of IUC Substance:

Type	Name	EC No.	CAS No.	Typical concentration	Concentration ranges			Structure
Core	Zno	215-222-5	1314-13-2	100 µg/ml	0 % (w/w)	0 % (w/w)	Also contained in...	Zn=O

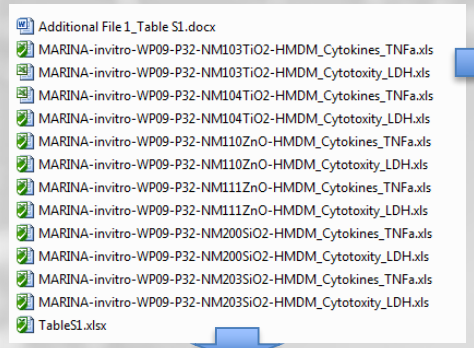
Search:

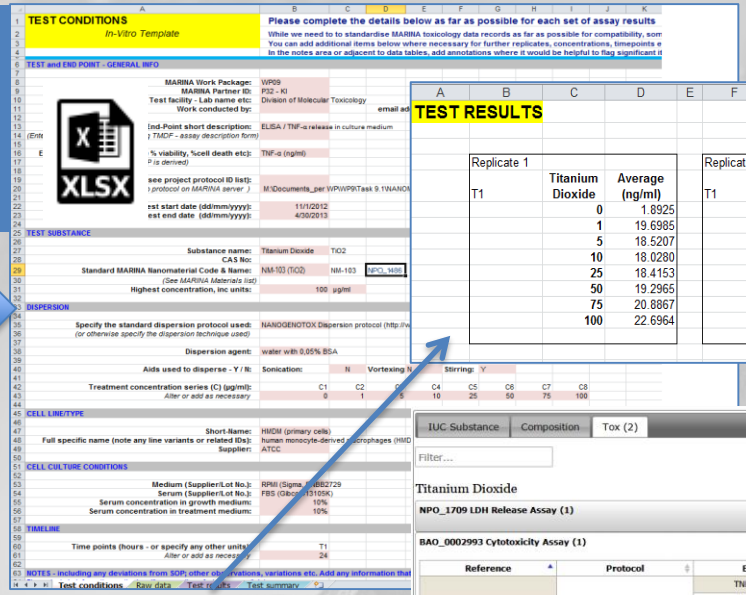
- 5 -	NM-200 (SiO2)	XLSX-93b2da85...	NPO_1373	Silica Dioxide	XLSX-93b2da85...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-200
- 6 -	NM-203 (SiO2)	XLSX-aacdfad8-...	NPO_1373	Silica Dioxide	XLSX-aacdfad8-...	FP7 MARINA	JRC Representative Manufactured Nanomaterials = NM-203

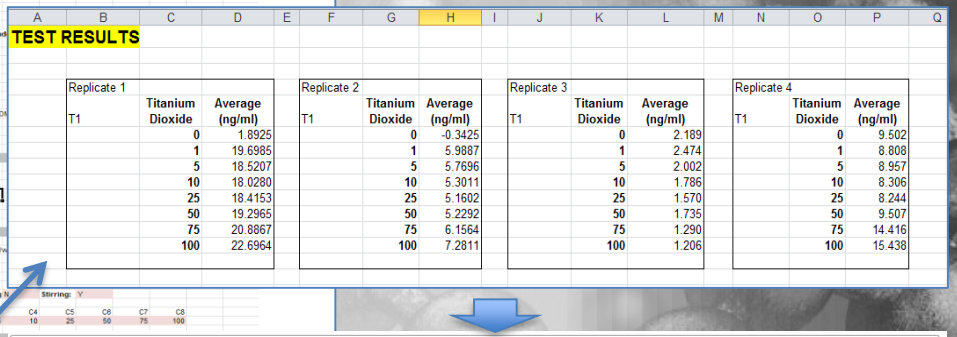
eNM eNanoMapper

From spreadsheets to DB content

FP7 MARINA /KI
doi:10.1371/journal.pone.0127174
Farcal et al. 2015

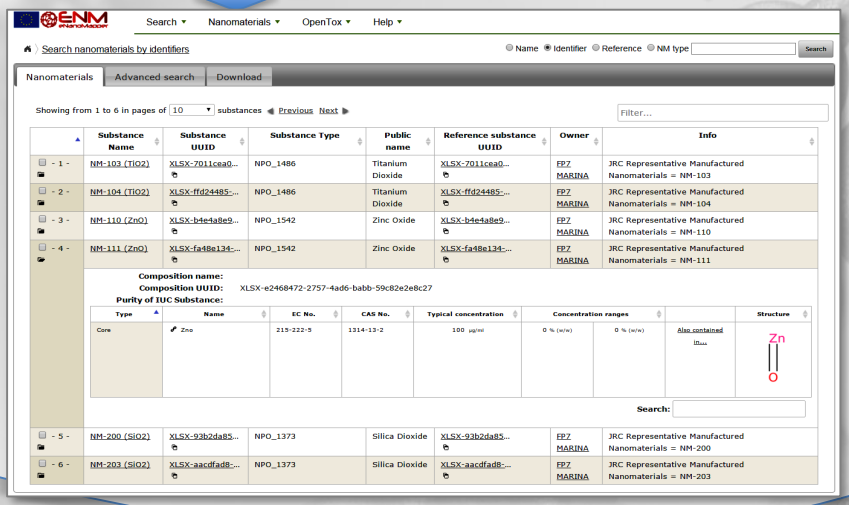


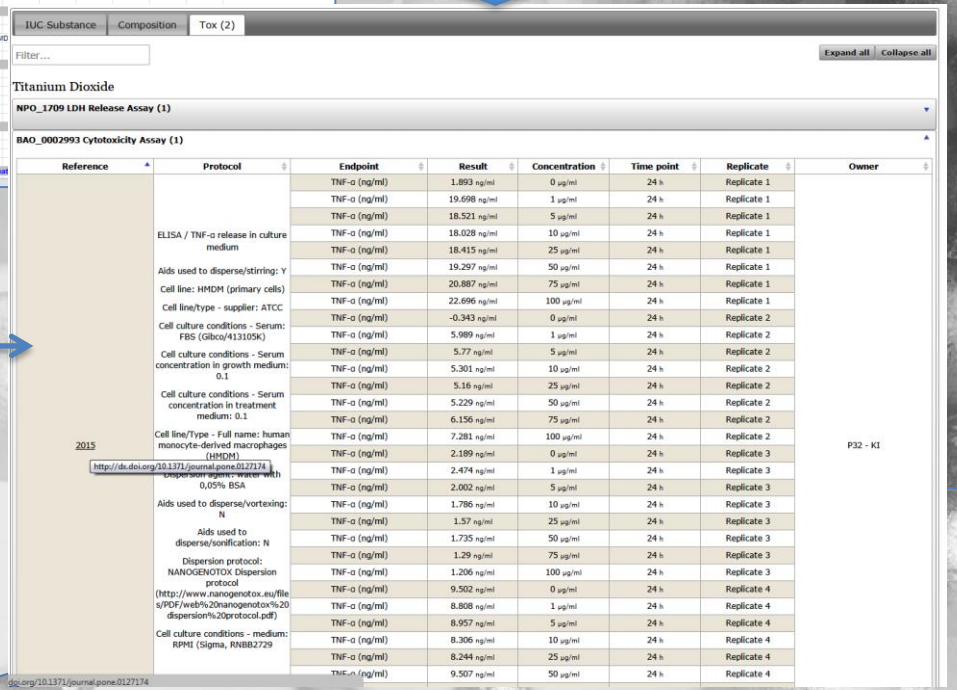




TEST RESULTS

Replicate 1			Replicate 2			Replicate 3			Replicate 4		
T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)
	0	1.8925		0	-0.3425		0	2.189		0	9.502
	1	19.6985		1	5.9887		1	2.474		1	8.808
	5	18.5207		5	5.7696		5	2.002		5	8.957
	10	18.0280		10	5.3011		10	1.786		10	8.306
	25	18.4153		25	5.1602		25	1.570		25	8.244
	50	19.2965		50	5.2292		50	1.735		50	9.507
	75	20.8867		75	6.1564		75	1.290		75	14.416
	100	22.6964		100	7.2811		100	1.206		100	15.438



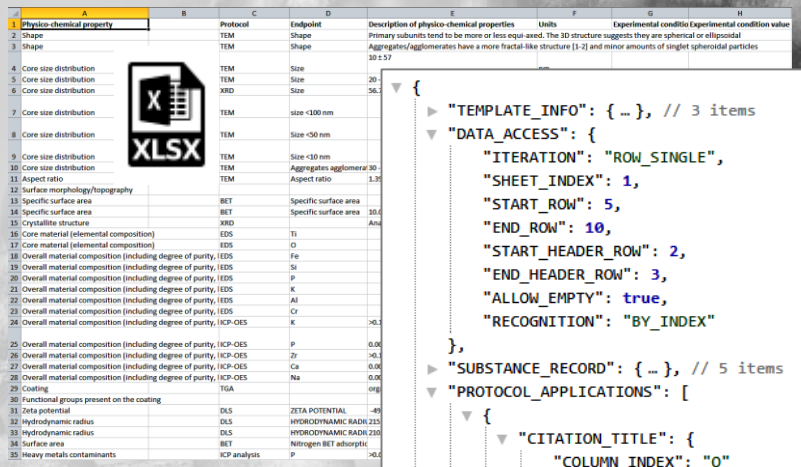


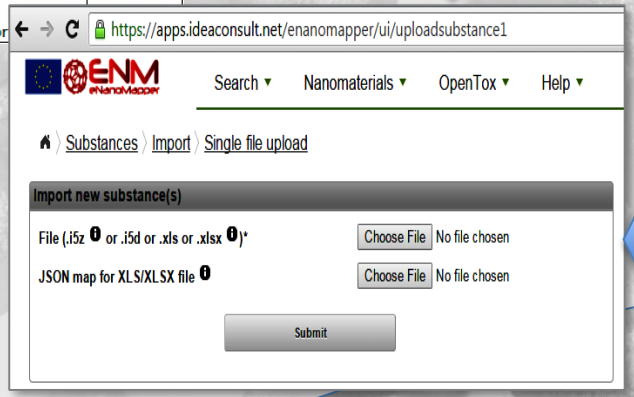
eNanoMapper

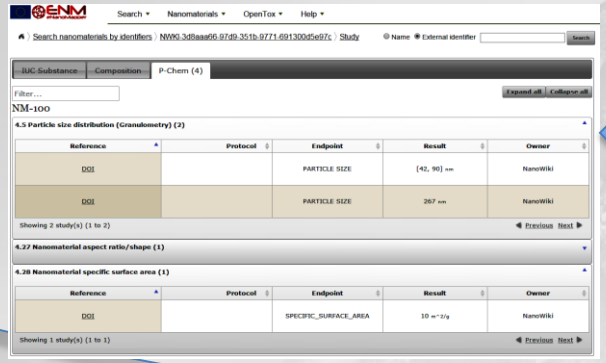
Data sources 2. PDF files

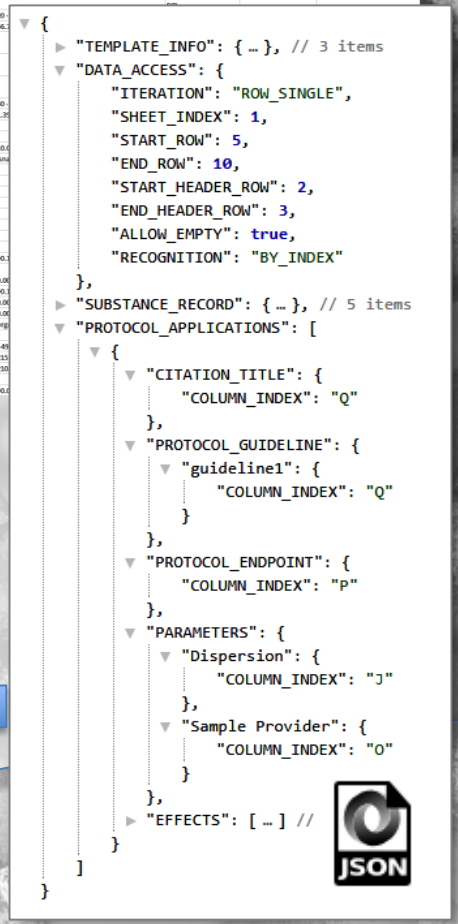
Summary of physico-chemical data			
Physico-chemical property (if applicable)	**Description of physico-chemical properties** (if applicable)	**Source of information**	**Date**
Shape	**TEM:** Primary subunits tend to be more or less equi-axed. The 3D structure suggests they are spherical or ellipsoidal **[2]** Aggregates/agglomerates have a more fractal-like structure **[1-2]** and minor amounts of singlet spheroidal particles **[1]**	**[1]** NANOGENOTOX: NRCWE data **[2]** JRC technical report	28.01.14
Core size distribution	**TEM:** 110 ± 57 nm **[1]**; from 20 to	**[1]** NANOGENOTOX:	28.01.14
	BET: 10.03 m2/g (room temp) **[2]**	Deliverable 4.4 **[2]** JRC technical repor	

Phys-chem characterisation of representative materials
JRC Nanomaterials Repository
Juan Riego-Sintes
https://ec.europa.eu/jrc/en/scientific-tool/jrc-nanomaterials-repository







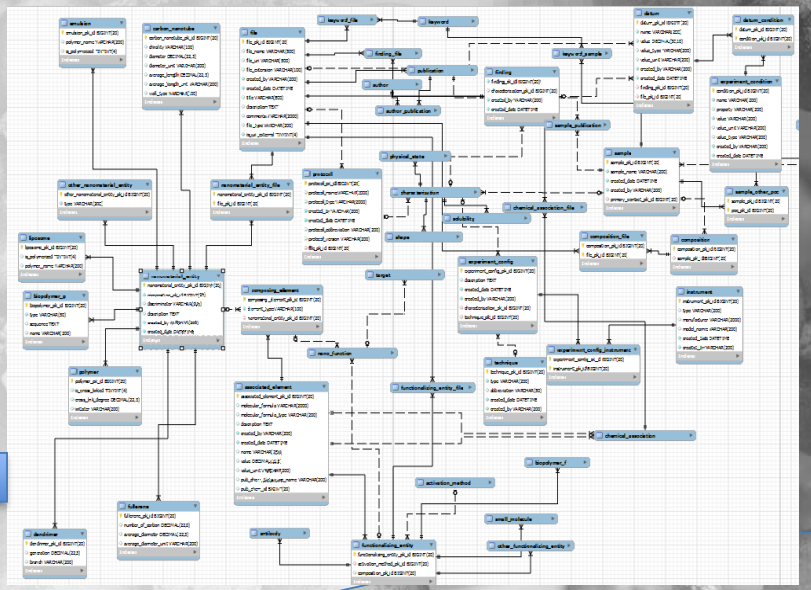


Data sources 3. SQL database (e.g. caNanoLab)

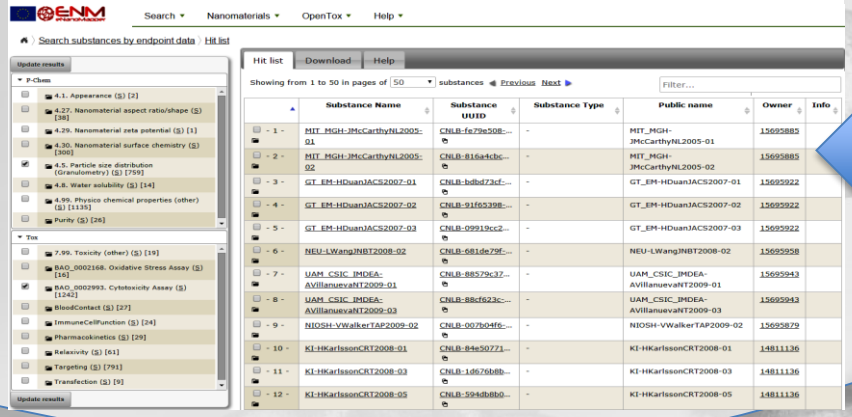
- MySQL dump provided to eNanoMapper
- Domain model class names and attributes https://cdebrowser.nci.nih.gov/CDEBrowser/
- Definitions for caNanoLab concepts are maintained in the NCI Enterprise Vocabulary Services https://ncit.nci.nih.gov/ncitbrowser
- Code: https://github.com/NCIP/cananolab

National Cancer Institute

caNanoLab



Mapping into eNanoMapper data model



eNM eNanoMapper

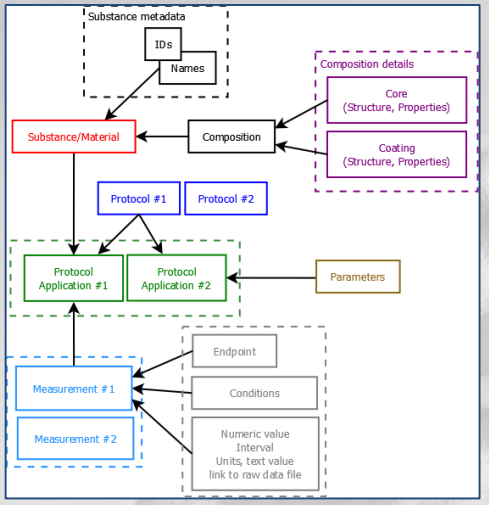
eNanoMapper data format conversion

Generic Excel Parser

...

OECD HT

ISA-TAB



JSON

ISA-TAB

ISA-JSON

...

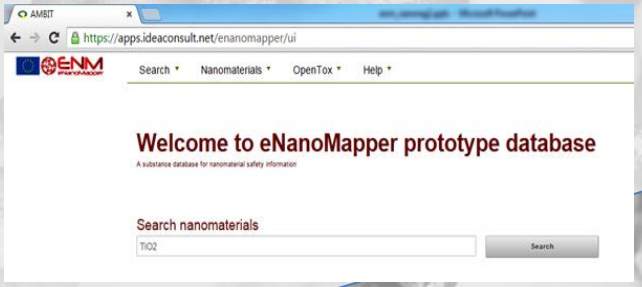
RDF

isa-api

BAO , linked ISA, ...

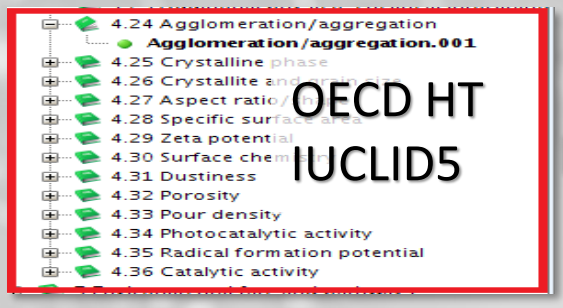
W3C SPARQL

Yet another search service

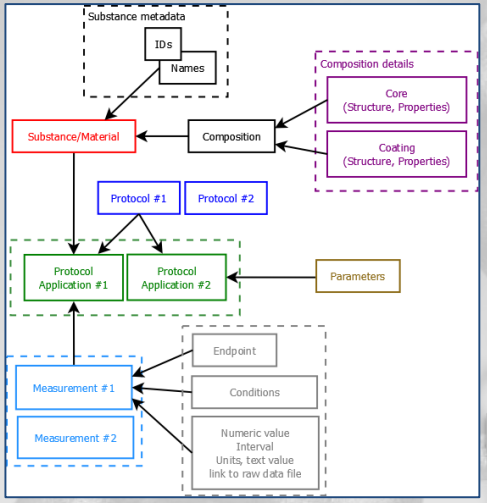


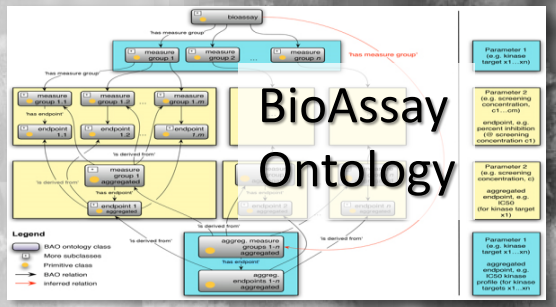
eNM eNanoMapper

eNanoMapper supported data formats

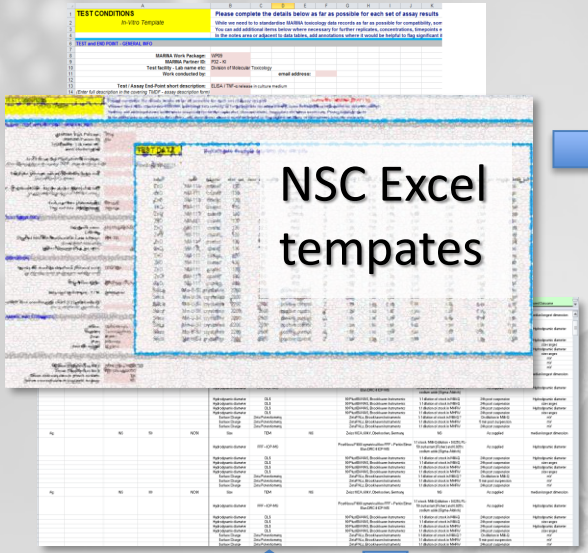


OECD HT
IUCLID5



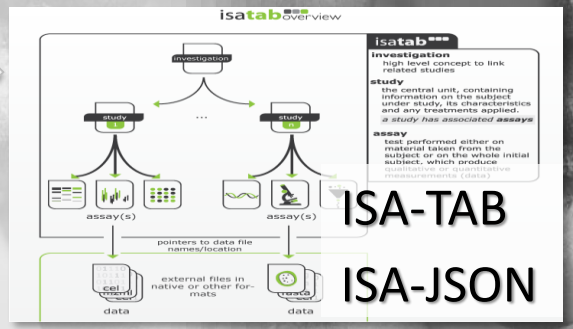


BioAssay
Ontology

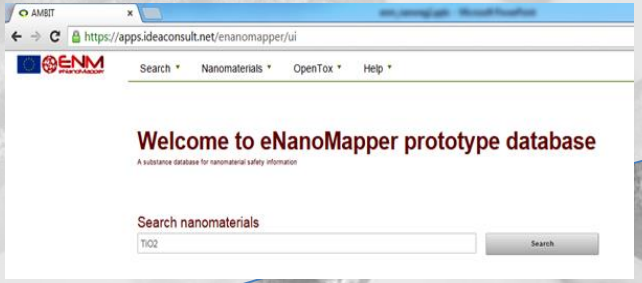


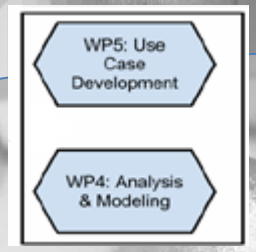
NSC Excel
tempates

REST API
JSON



ISA-TAB
ISA-JSON





WP2: Ontology Development

isatab

ISA-TAB: Susanna-Assunta Sansone et. al., **Toward interoperable bioscience data**, Nature Genetics 44, 121–126 (2012)

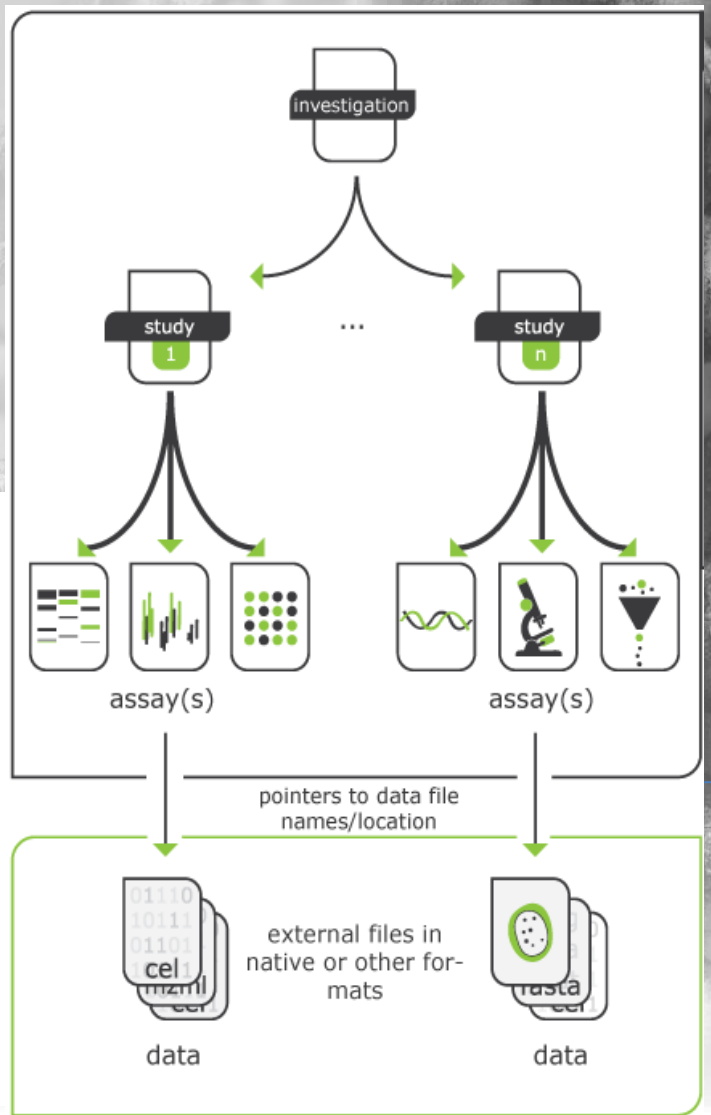
- Version 1 (Nov 2008)
- Data is described in 3 layers
- Tab delimited format (*.txt and *.xls/xlsx)
- Only meta data is stored
- Pointers to the data files
- Ontology references
- Additional configurations
- **Version 2 (2015)**
- **(under development)**
- **ISA-JSON (2015)**
- **(under development)**

investigation
high level concept to link related studies

study
the central unit, containing information on the subject under study, its characteristics and any treatments applied.

a study has associated ***assays***

assay
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)



isatab (Investigation file)

	A	B
1	ONTOLOGY SOURCE REFERENCE	
2	Term Source Name	CHEBI
3	Term Source File	http://data.bioontology.org/ontologies/CHEBI
4	Term Source Version	78
5	Term Source Description	Chemical Entities of Biological Interest Ontology
6	INVESTIGATION	
7	Investigation Identifier	BII-I-1
8	Investigation Title	Growth control of the eukaryote cell: a systems biology study in yeast
9	Investigation Description	Background Cell growth underlies many key cellular and developmen
10	Investigation Submission Date	4/30/2007
11	Investigation Public Release Date	3/10/2009
12	Comment [Created with configuration]	
13	Comment [Last Opened With Configuration]	isaconfig-default_v2013-02-13
14	Comment [Owning Organisation URI]	
15	Comment [Consortium URI]	
16	Comment [Principal Investigator URI]	
17	Comment [Investigation keywords]	
18	INVESTIGATION PUBLICATIONS	
19	Investigation PubMed ID	17439666
20	Investigation Publication DOI	doi:10.1186/jbiol54
21	Investigation Publication Author List	Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell M
22	Investigation Publication Title	Growth control of the eukaryote cell: a systems biology study in yeast
23	Investigation Publication Status	indexed in Pubmed
24	Investigation Publication Status Term Accession Number	

isatab (study and assay files)

Investigation consists of one or more studies

Study Person Roles Term Source REF	
Comment[Study Person REF]	
STUDY	
Study Identifier	BII-S-2
Study Title	A time course analysis of transcription response in yeast treated with r
Study Description	Comprehensive high-throughput analyses at the levels of mRNAs, prot
Comment[Study Grant Number]	
Comment[Study Funding Agency]	
Study Submission Date	4/30/2007
Study Public Release Date	3/10/2009
Study File Name	s_BII-S-2.txt
STUDY DESIGN DESCRIPTORS	
Study Design Type	time series design
Study Design Type Term Accession Number	http://purl.obolibrary.org/obo/OBI_0500020
Study Design Type Term Source REF	OBI
STUDY PUBLICATIONS	

pointer to the study file

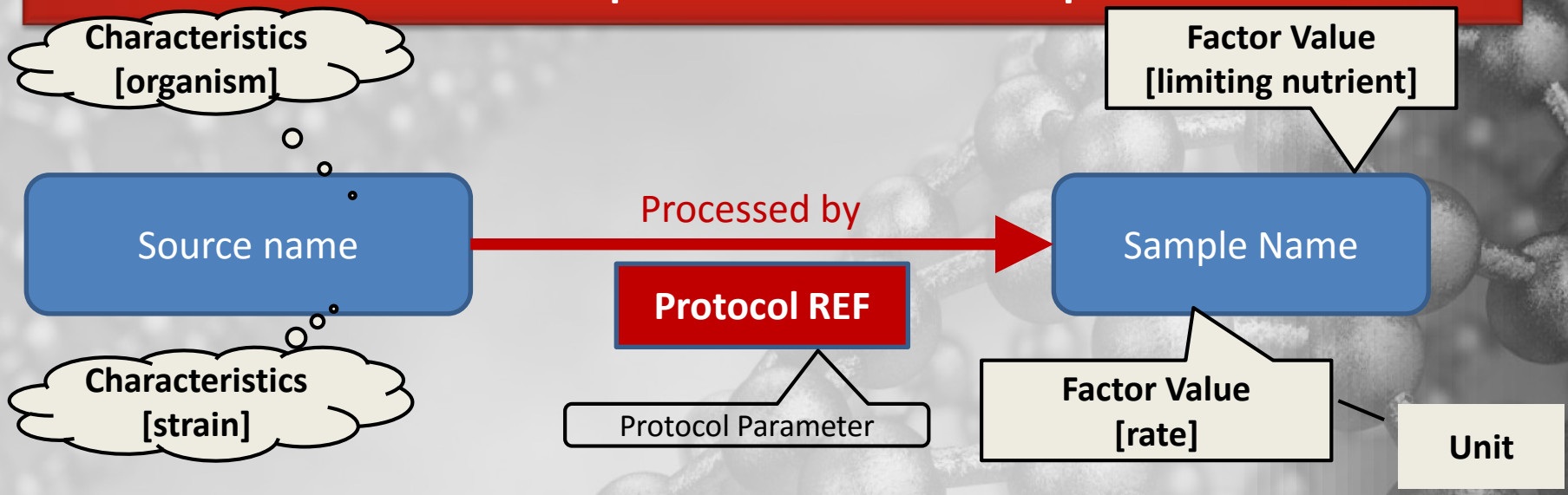
Each study contains one or more assays

pointers to the assay files

STUDY ASSAYS			
Study Assay File Name	a_metabolome.txt	a_proteome.txt	a_transcriptome.txt
Study Assay Measurement Type	metabolite profiling	protein expression profiling	transcription profilin
Study Assay Measurement Type Term Accession Number	http://purl.obolibrary.org/obo/OBI_00	http://purl.obolibrary.org/o	42

eNanoMapper

isatab data model: The Experimental Graph



Source Name	Characteristics [organism]	Characteristics [strain]	Protocol REF	Sample Name	Factor Value [limiting nutrient]	Factor Value [rate]	Unit
culture1	Saccharomyces cerevisiae	FY1679	growth protocol	C-0.07-aliquot1	carbon	0.07	l/hour
culture4	Saccharomyces cerevisiae	FY1679	growth protocol	N-0.07-aliquot1	nitrogen	0.07	l/hour
culture5	Saccharomyces cerevisiae	FY1679	growth protocol	N-0.1-aliquot1	nitrogen	0.1	l/hour

ISA-TAB-Nano

The ISA-TAB-Nano project is an effort of the National Cancer Institute (**NCI**) National Cancer Informatics Program (**NCIP**) Nanotechnology Informatics Working Group (**Nano WG**)

NIH NATIONAL CANCER INSTITUTE
NCI Wiki

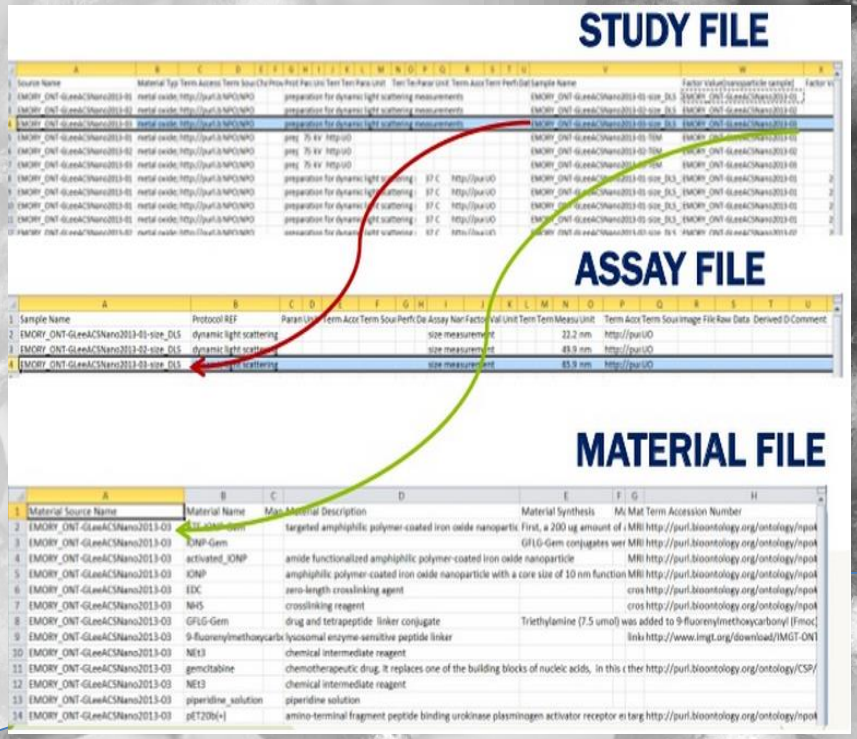
- **ISA-TAB-Nano site**

https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano

- ISA-TAB-Nano templates to create ISA-TAB-Nano files
- Template glossary for definitions
- Example files

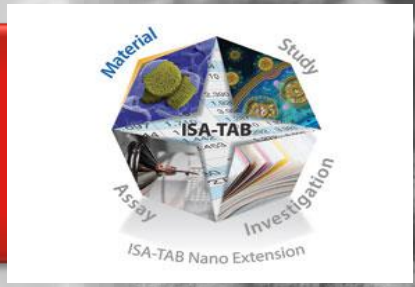
- **Publications**

- BMC Biotechnology 2013, 13:2
  http://www.biomedcentral.com/1472-6750/13/2/abstract
- Commentary Nature Nanotechnology 2013, 8, 73-74
  http://www.nature.com/nnano/journal/v8/n2/full/nnano.2013.12.html



eNM eNanoMapper

ISA-TAB-Nano



Material M is comprised of A, B, and C

Material Source Name	Material Name	Material Constituent	Material Linkage Type
M-lab-1	M	A;B;C	
M-lab-1	M	A;B	covalent linkage
M-lab-1	M	A; C	covalent linkage
M-lab-1	A		
M-lab-1	B		
M-lab-1	C		

Material M is comprised of A, B, and C within Material M, A and B are linked and A and C are linked

eNM eNanoMapper

ISA-JSON project
https://github.com/ISA-tools/isa-api

- Under development by Oxford group and collaborators
- Python based **ISA** API library
- New data format based on JSON describes the ISA experimental graph
  - More efficient data storage than the TAB delimited
- Full support of the old ISA-TAB format (v.1)
- New extended ISA v.2

- eNM – ISA-Tab team meeting May 2015 (EBI)
- ISA-Tab hackathon July 2015
- eNM- ISA-Tab team virtual meeting Nov 3 2015
- ISA Working group teleconference Dec 9 2015
- Active collaboration on github
  - schema, issues, code

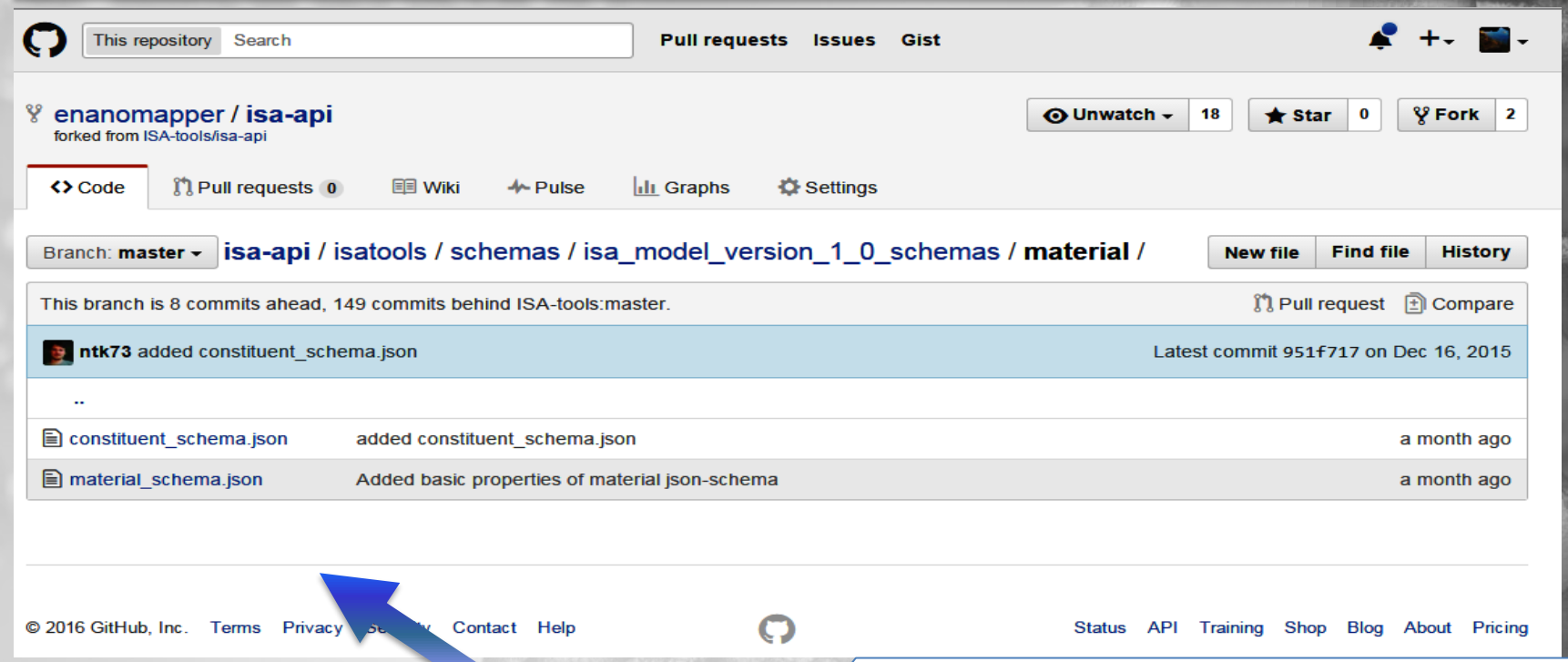
ISA-JSON schemas

https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa_model_version_1_0_schemas/core

- assay_schema.json
- comment_schema.json
- data_schema.json
- factor_schema.json
- factor_value_schema.json
- investigation_schema.json
- material_attribute_schema.json
- material_schema.json
- ontology_annotation_schema.json
- ontology_source_reference_schema.json
- organization_schema.json
- parameter_schema.json
- person_schema.json
- process_schema.json
- protocol_schema.json

```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA investigation schema",
    "description" : "JSON-schema representing an investigation in the ISA model",
    "type" : "object",
    "properties" : {
        "identifier" : { "type" : "string" },
        "title" : { "type" : "string"},
        "description" : { "type" : "string"},
        "submissionDate" : { "type" : "string", "format" : "date-time"},
        "publicReleaseDate" : { "type" : "string", "format" : "date-time"},
        "commentCreatedWithConfiguration" : {
            "$ref": "comment_schema.json#",
            "name": "Created With Configuration"
        },
        "commentLastOpenedWithConfiguration" : {
            "$ref": "comment_schema.json#",
            "name": "Last Opened With Configuration"
        },
        "ontologySourceReferences" : {
            "type" : "array",
            "items" : {
                "$ref": "ontology_source_reference_schema.json#"
            }
        },
```

eNanoMapper: ISA-JSON material extension



Contributing new extension to isa-api (under development by eNM team)

JSON schema corresponding to ISA-TAB-Nano material file

eNanoMapper:
ISA (v.1) Java classes

- Assay.java
- Comment.java
- Component.java
- Data.java
- Factor.java
- FactorValue.java
- Investigation.java
- Material.java
- MaterialAttribute.java
- MeasurementType.java
- OntologyAnnotation.java
- OntologySourceReference.java
- Organization.java
- Parameter.java
- Person.java
- Process.java
- Protocol.java
- Publication.java
- Role.java
- Sample.java
- Source.java
- Study.java
- TechnologyType.java
- Value.java

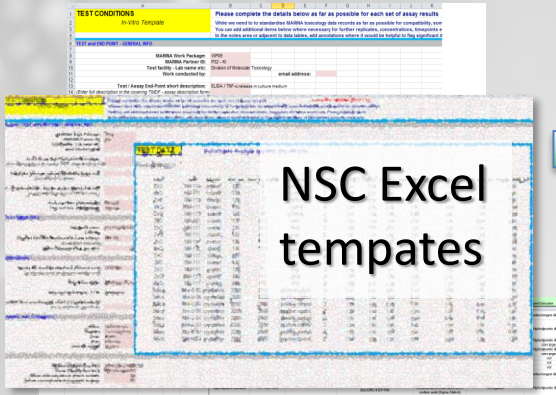
- Investigation.java
  - Investigation
    - commentCreatedWithConfiguration
    - commentLastOpenedWithConfiguration
    - description
    - identifier
    - ontologySourceReferences
    - people
    - publications
    - publicReleaseDate
    - studies
    - submissionDate
    - title

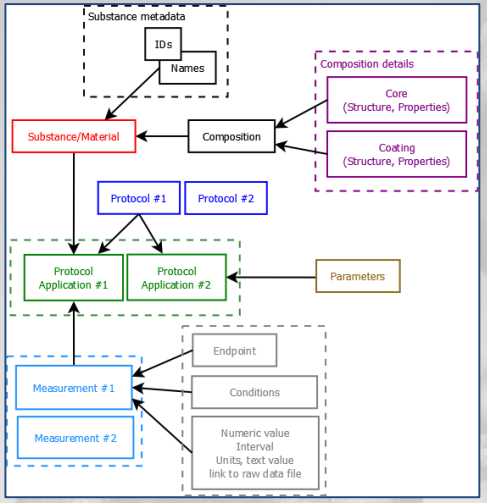
```
    @JsonProperty("identifier")
    public String identifier;
    @JsonProperty("title")
    public String title;
    @JsonProperty("description")
    public String description;
    @JsonProperty("submissionDate")
    public Date submissionDate;
    @JsonProperty("publicReleaseDate")
    public Date publicReleaseDate;

    @JsonProperty("commentCreatedWithConfiguration")
    public Comment commentCreatedWithConfiguration;
    @JsonProperty("commentLastOpenedWithConfiguration")
    public Comment commentLastOpenedWithConfiguration;
    @JsonProperty("ontologySourceReferences")
    public List<OntologySourceReference> ontologySourceReferences = new A
    @JsonProperty("publications")
    public List<Publication> publications = new ArrayList<Publication>();
    @JsonProperty("people")
    public List<Person> people = new ArrayList<Person>();
    @JsonProperty("studies")
    public List<Study> studies = new ArrayList<Study>();
}
```

- Study.java
  - Study
    - assays
    - description
    - identifier
    - people
    - processSequence
    - protocols
    - publications
    - publicReleaseDate
    - samples
    - sources
    - studyDesignDescriptors
    - submissionDate
    - title

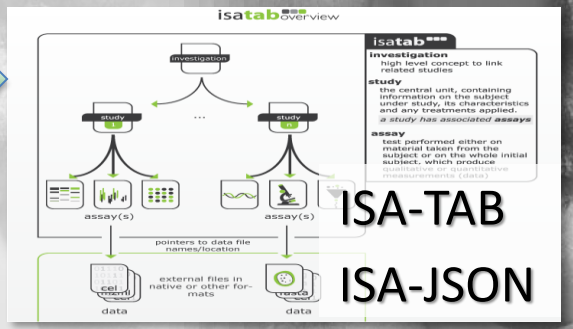
```
public class Study {

    @JsonProperty("identifier")
    public String identifier;
    @JsonProperty("title")
    public String title;
    @JsonProperty("description")
    public String description;
    @JsonProperty("submissionDate")
    public Date submissionDate;
    @JsonProperty("publicReleaseDate")
    public Date publicReleaseDate;
    @JsonProperty("publications")
    public List<Publication> publications = new ArrayList<Publication>();
    @JsonProperty("people")
    public List<Person> people = new ArrayList<Person>();
    @JsonProperty("studyDesignDescriptors")
    public List<OntologyAnnotation> studyDesignDescriptors = new ArrayList<Ontolog
    @JsonProperty("protocols")
    public List<Protocol> protocols = new ArrayList<Protocol>();
```

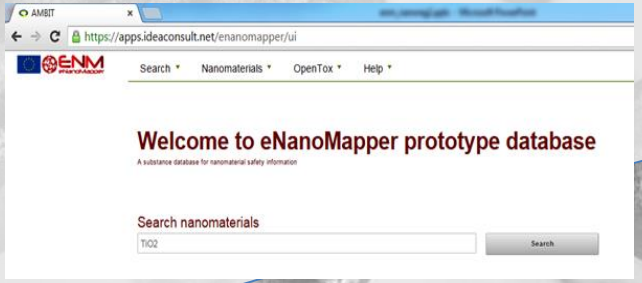
Automatic conversion to/from ISA-JSON/ISA-TAB (under development)





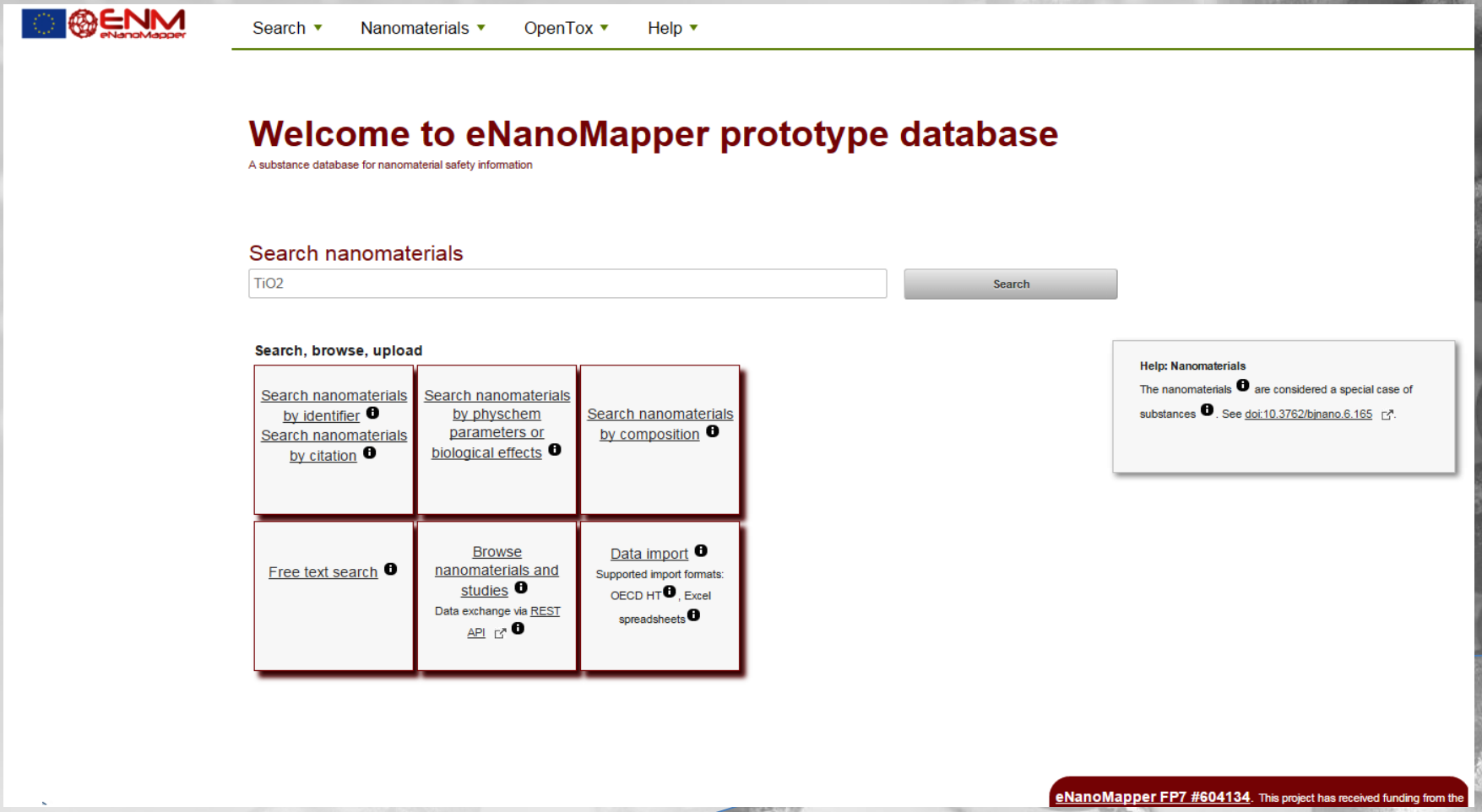
REST API
JSON



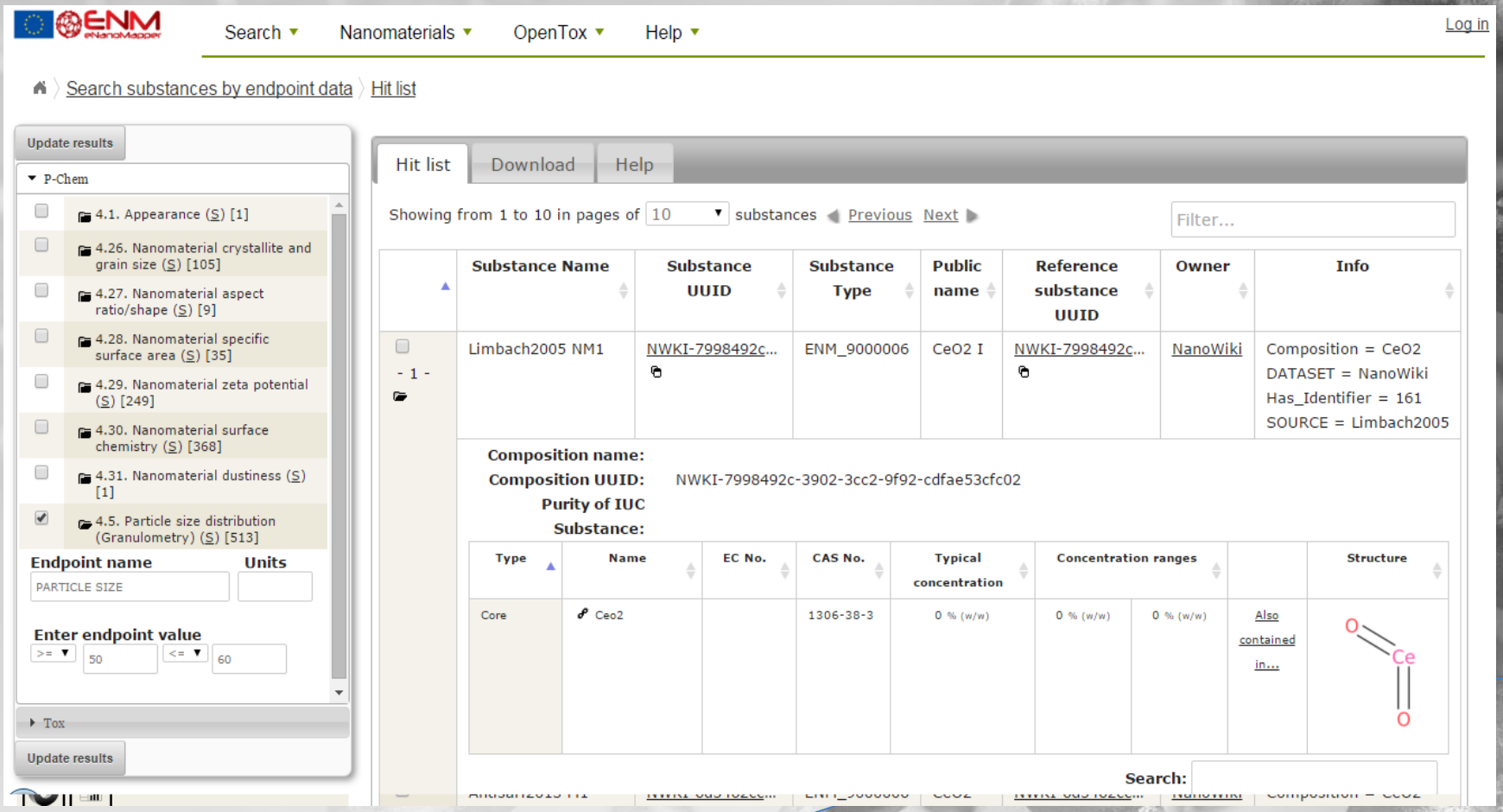


WP2:
Ontology
Development

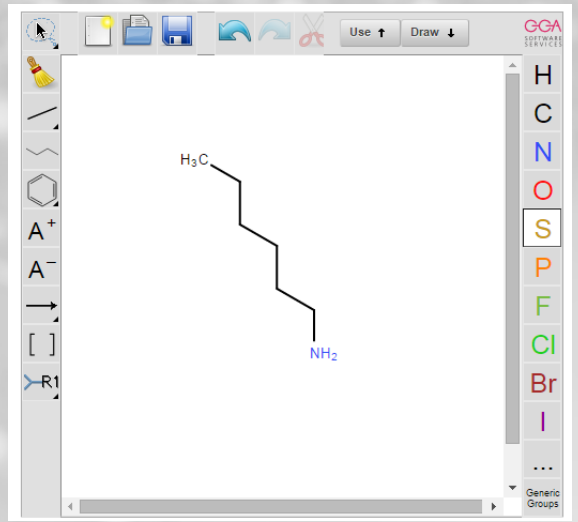
Searchability



Search by physchem and bio effects

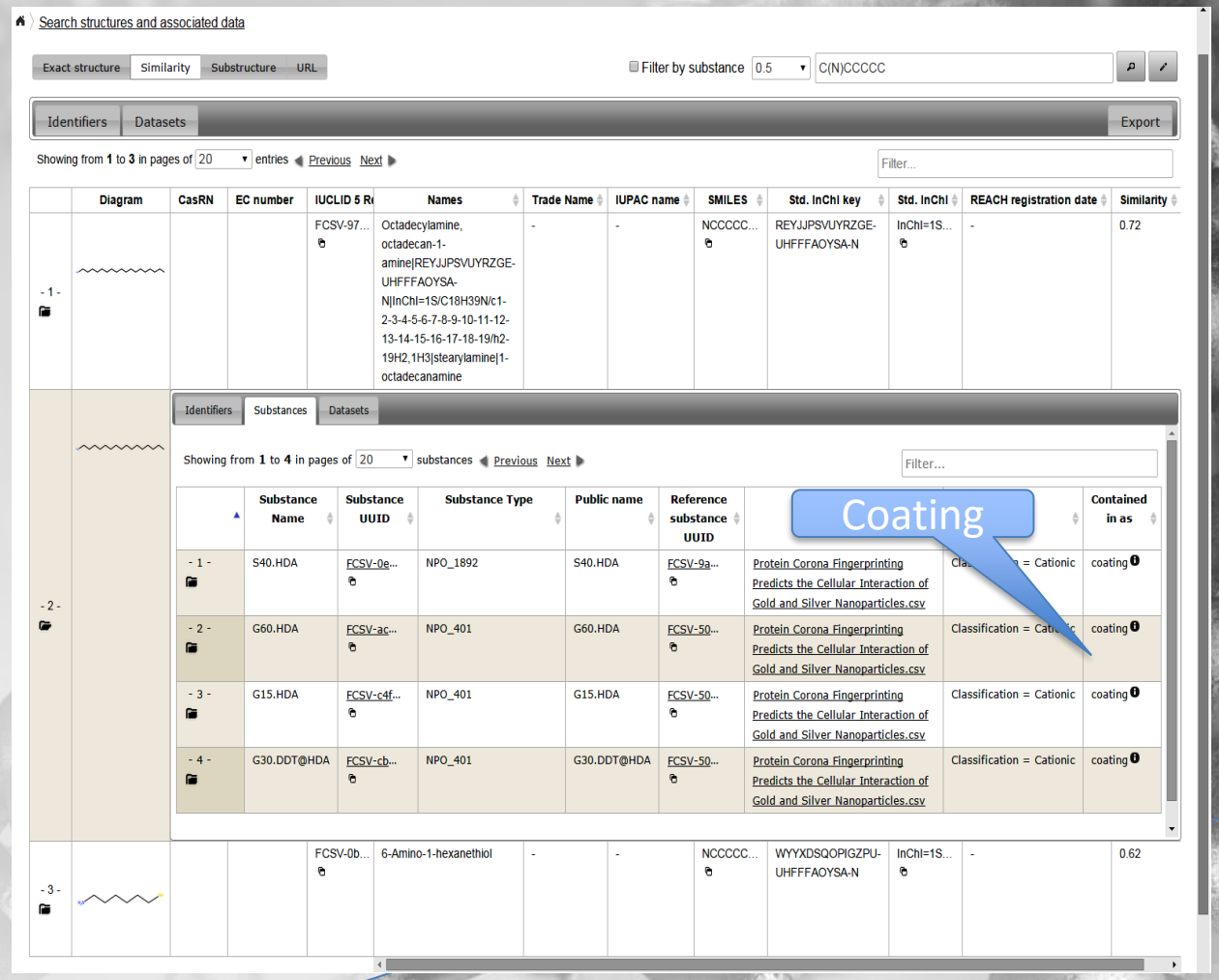


Search by chemical composition



Chemical similarity is a pivotal concept in cheminformatics, encompassing a variety of computational methods quantifying the extent to which two chemical structures resemble each other.

There is not yet a standardized approach for NM similarity, however a number of attempts for NM grouping and read across have been published recently.

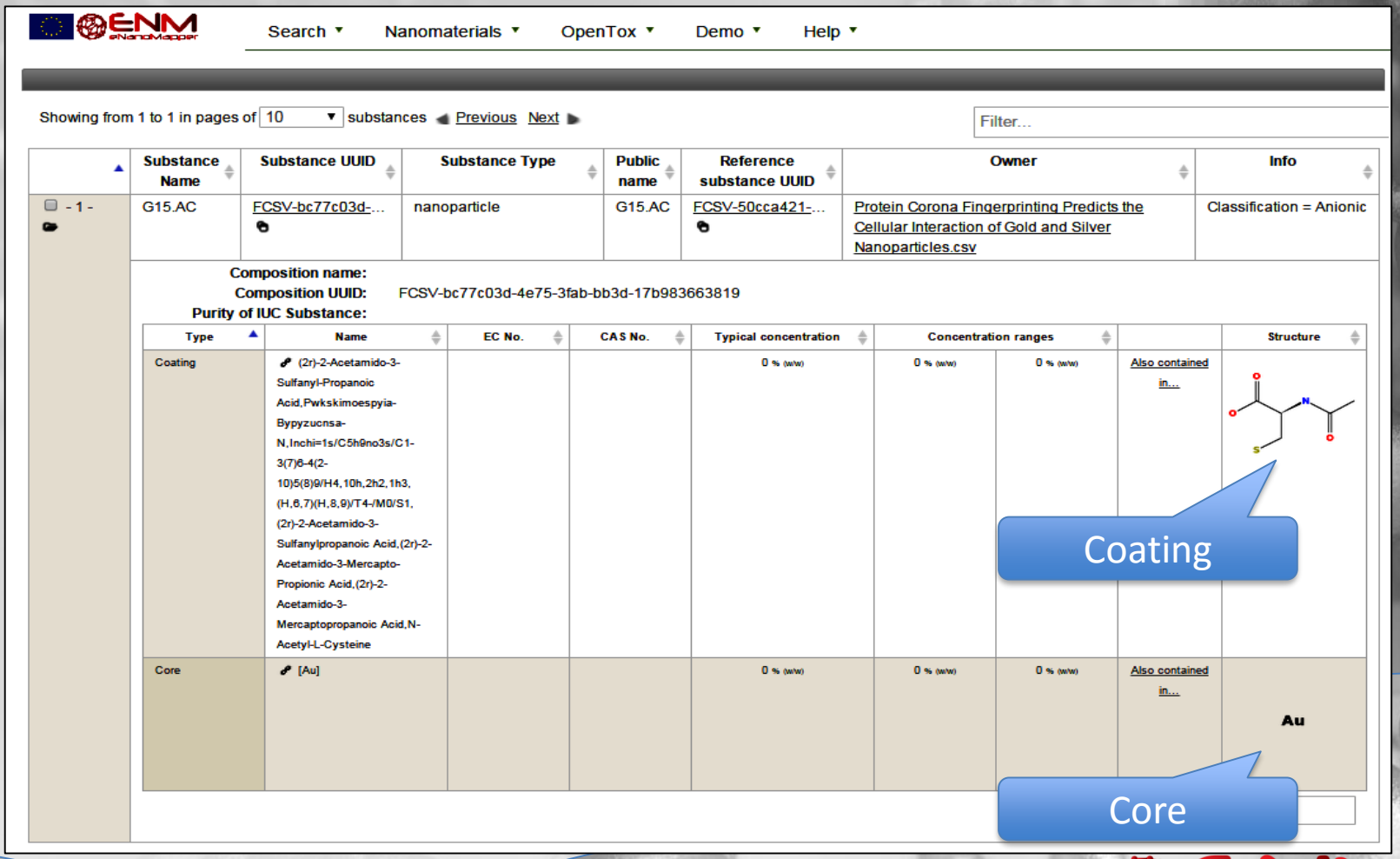


eNanoMapper

Substances / Materials representation

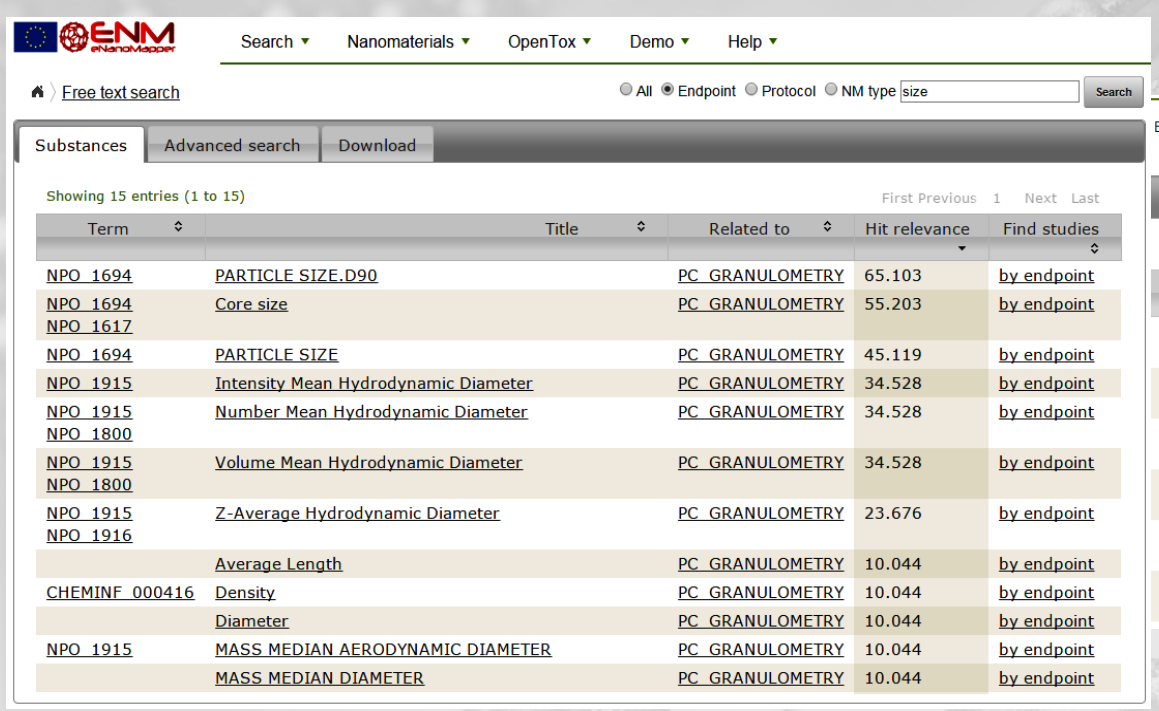
- ***NanoParticle Ontology (NPO):*** a **Nanomaterial (NPO_199) is an equivalent class to chemical substance (NPO_1973)** one of (nano-object, nanoparticle, engineered nanomaterial, nanostructured material, nanoparticle formulation). The chemical substance itself is a subclass of a chemical entity (NPO_1972).
- **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.
- **IUPAC: "Matter of constant composition best characterized by the entities (molecules, formula units, atoms) it is composed of. Physical properties such as density, refractive index, electric conductivity, melting point etc. characterize the chemical substance"**

Material composition



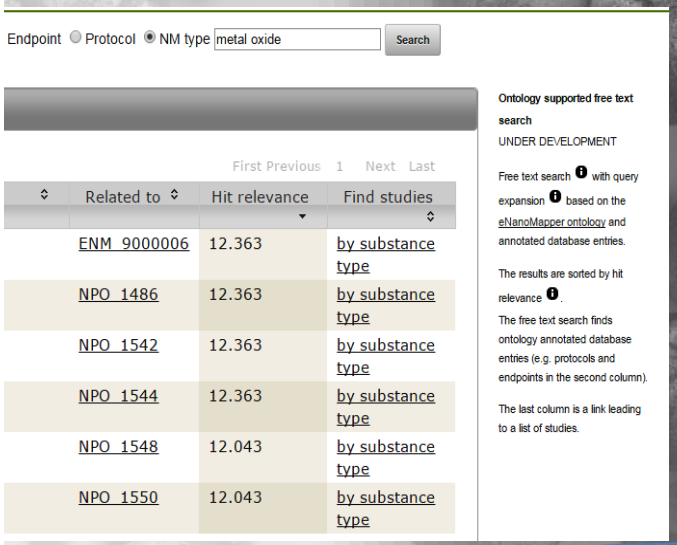
Free text search supported by ontology annotated database

- Several technical options available



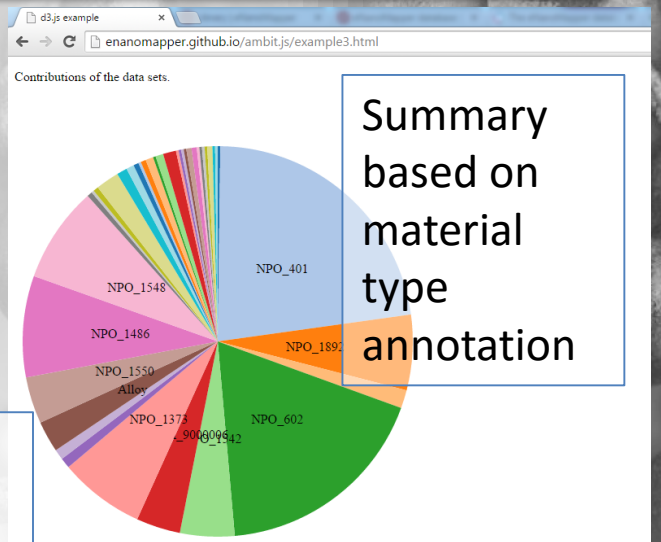
Search for "size"

Search for "metal oxide"



The search quality depends on ontology annotation

- How to annotate
  - Data model supported (minimal)
  - Manual annotation
  - Parser supported annotation
  - Automated heuristics
- Entries to be annotated
  - Material types
  - Assays
  - Endpoints
  - Protocol parameters
  - Experiment factors
  - Experiment results
  - Units
  - NM providers,
  - NM identifiers, etc.



Summary based on material type annotation

Summary based on assay annotation

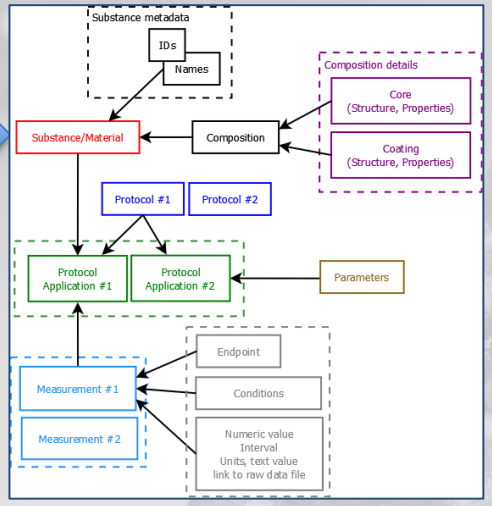
Row Labels	P-CHEM	TOX
4.1. Appearance	1	
4.26. Nanomaterial crystallite and grain size	105	
4.27. Nanomaterial aspect ratio/shape	9	
4.28. Nanomaterial specific surface area	35	
4.29. Nanomaterial zeta potential	249	
4.30. Nanomaterial surface chemistry	368	
4.31. Nanomaterial dustiness	1	
4.5. Particle size distribution (Granulometry)	513	
7.2.1. Acute toxicity - oral		1
7.6.1. Genetic toxicity in vitro		1
7.7. Carcinogenicity		1
7.99. Toxicity (other)		471
8.100. Proteomics		121
BAO_0002167. Genotoxicity Assay		12
BAO_0002993. Cytotoxicity Assay		20
BAO_0003009. Cell Viability Assay		200
NPO_1709. LDH Release Assay		6
Grand Total	1281	833

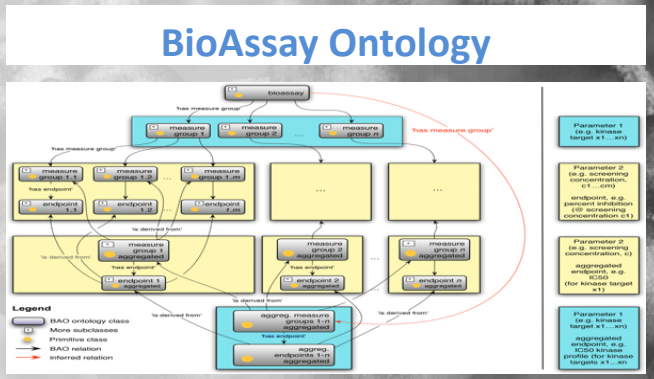
Free text & faceted search (under development)

Comments welcome at https://github.com/enanomapper/data.enanomapper.net/issues/8

Semantic search

Multiple Input formats





Semantic formats
RDF/XML
N3
JSON LD

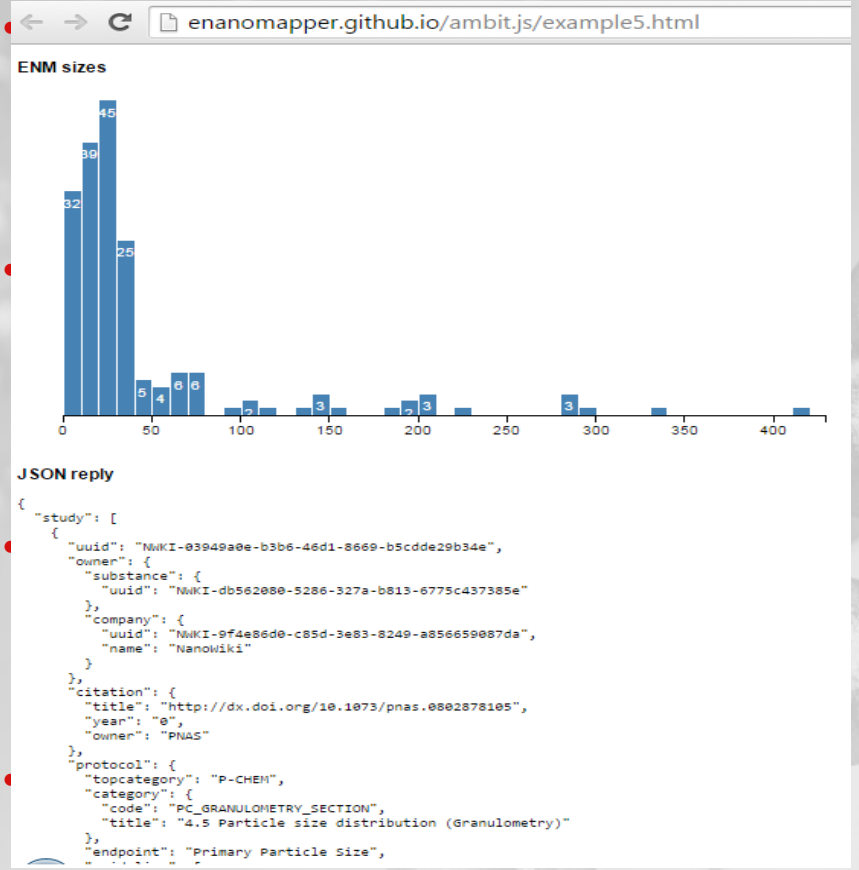
Triple store

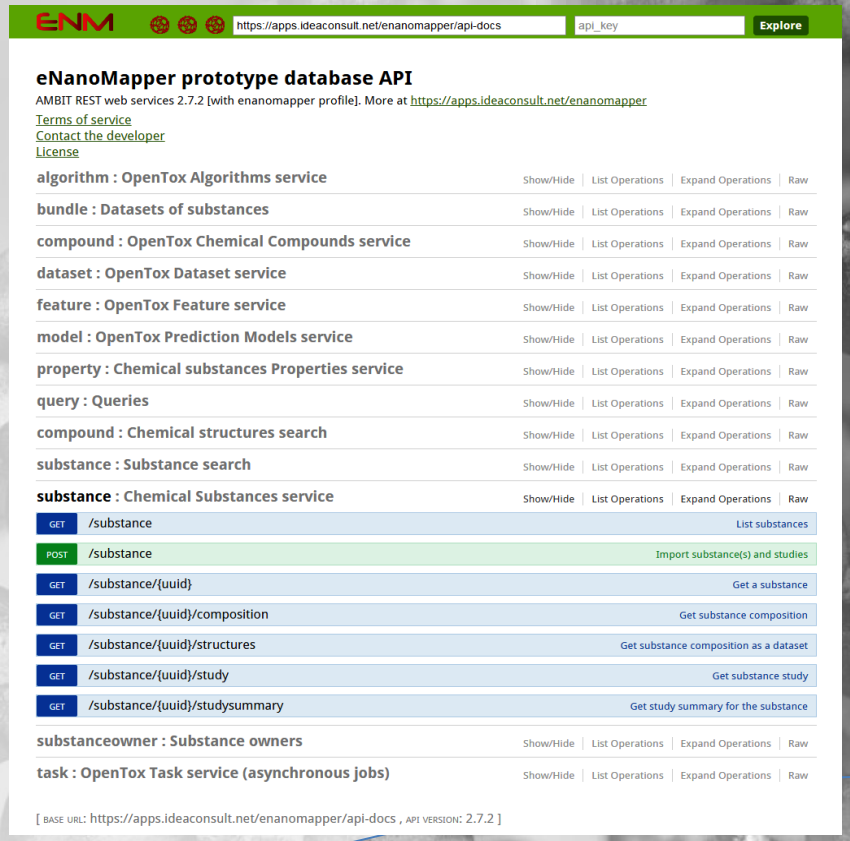
W3C SPARQL

Tailored interfaces

- Application Programming Interface (API)
- Use the API to
  - Develop user interface(s)
  - Communication with modelling tools
  - Integration with databases

eNanoMapper DB implementation : Application Programming Interface



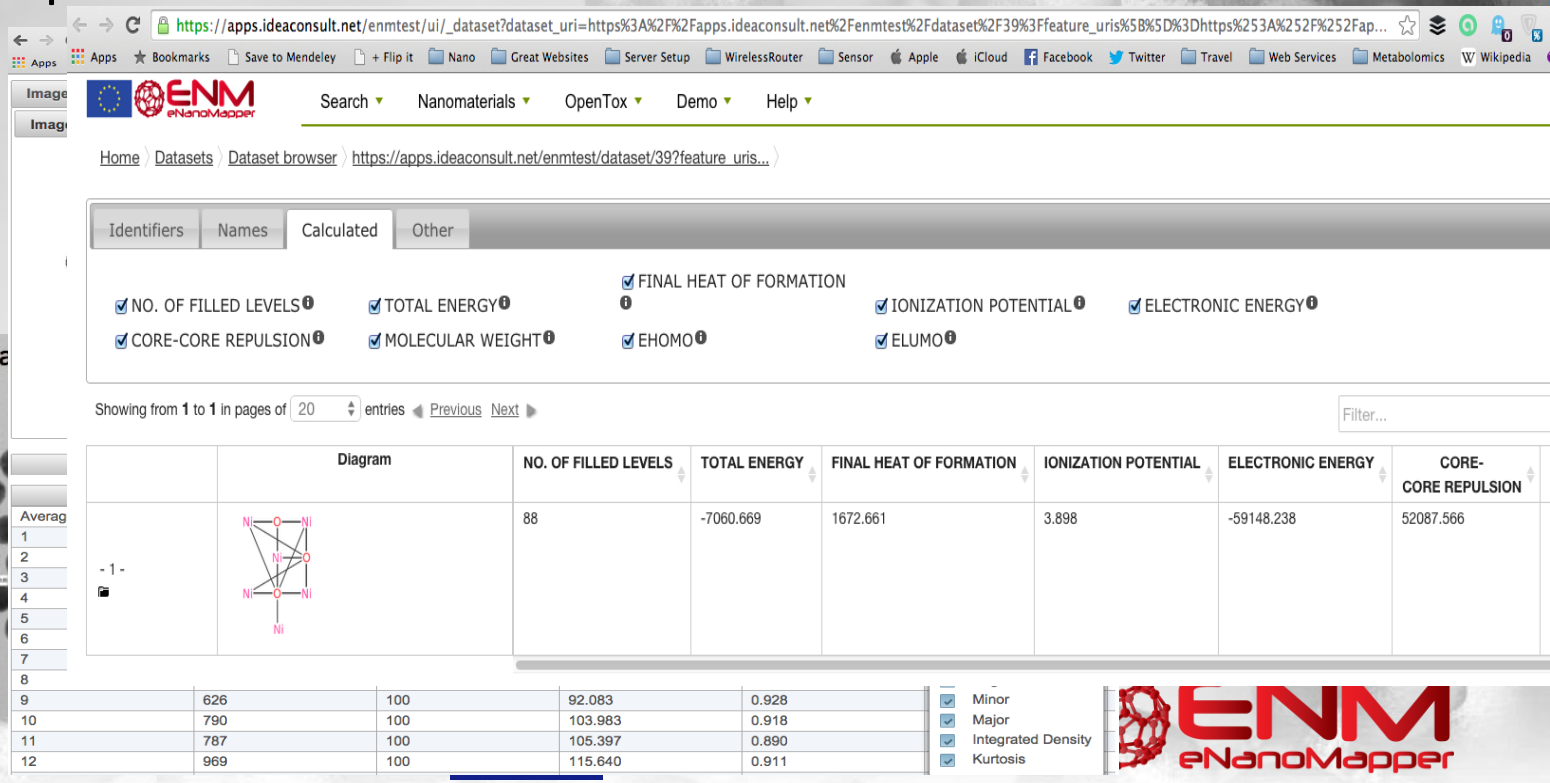


- **eNanoMapper database is based on the open source project http://ambit.sf.net**

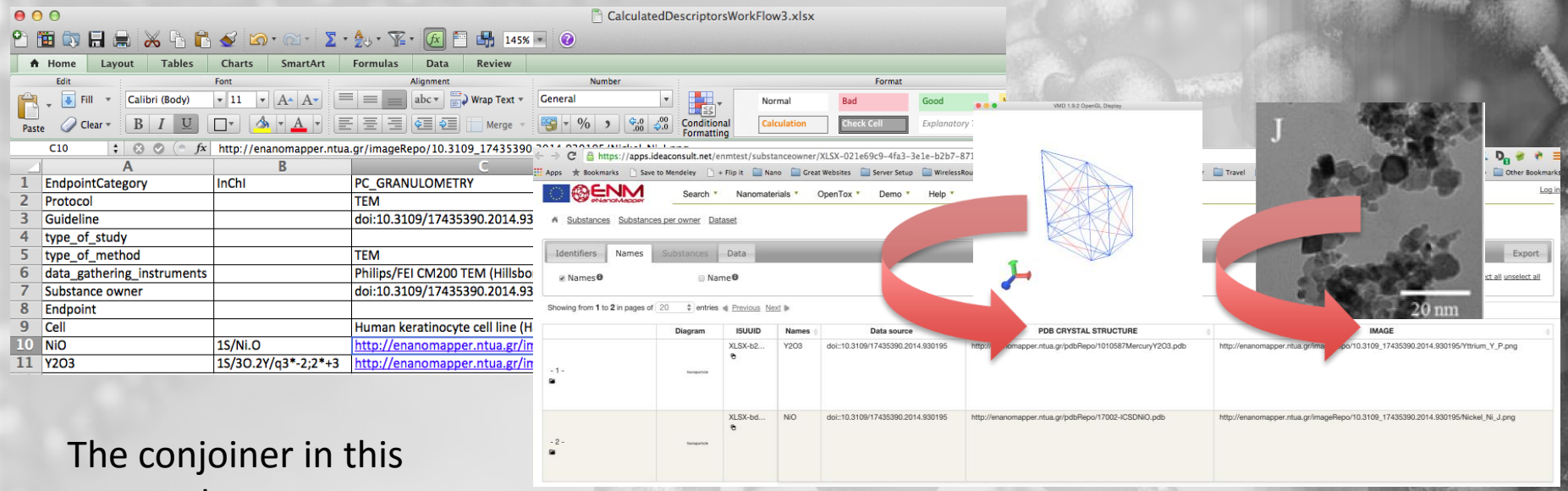
http://enanomapper.github.io/API/

WP4: Descriptor calculations (uses API to retrieve DB data and store back results)

- Image-derived Descriptors
- Quantum Mechanical Descriptors using MOPAC
- Gene Ontology Descriptors
- CDK Descriptors



WP4: A service to transform the data into ready-to-model-form



	A	B	C
1	EndpointCategory	InChI	PC_GRANULOMETRY
2	Protocol		TEM
3	Guideline		doi:10.3109/17435390.2014.93
4	type_of_study		
5	type_of_method		TEM
6	data_gathering_instruments		Philips/FEI CM200 TEM (Hillsbo
7	Substance owner		doi:10.3109/17435390.2014.93
8	Endpoint		
9	Cell		Human keratinocyte cell line (H
10	NiO	1S/Ni.O	http://enanomapper.ntua.gr/in
11	Y2O3	1S/3O.2Y/q3*-2;2*+3	http://enanomapper.ntua.gr/in

The conjoiner in this example converts heterogeneous data of metal oxides (image and crystal structure) into a dataset ready for modeling.

compound:
{
URI: "https://apps.ideaconsult.net/enmtest/substance/XLSX-b2a6a8e9-a7d4-349c-9fcc-df05356a508d"
},
values:
{
http://app.jaqpot.org:8080/jaqpot/services/feature/image+average+particle+angle: 149.23083,
http://app.jaqpot.org:8080/jaqpot/services/feature/image+average+particle+area: 8803,
http://app.jaqpot.org:8080/jaqpot/services/feature/image+average+particle+area_fraction: 43.979816,
http://app.jaqpot.org:8080/jaqpot/services/feature/image+average+particle+aspect_ratio: 1.3006951,
http://app.jaqpot.org:8080/jaqpot/services/feature/image+average+particle+circularity: 0.34530884,

eNM eNanoMapper

The need of integration … is not unique for nanomaterials

- 2005: “Integrated Informatics in Life and Materials Sciences: An Oxymoron?” *
  - Calculations, Descriptors, Statistics, Models
  - Data (substances, chemical structures, properties, predictions)
- Approaches toward integration:
  - Workflow management systems
  - Standalone container applications (chassis)
  - Web services, web mashups
  - Index –time integration vs query-time integration

** Gilardoni, F., Curcin, V., Karunanayake, K., Norgaard, J., & Guo, Y. (2005). QSAR Combinatorial Science, 24(1), 120-130.*

eNanoMapper

Chemical /Toxicogenomics DB
(no explicit NM support)

PubChem 10 Years

NM: Carbon nanotube assays
>200 fullerenes; metal oxides; silver nanoparticles; colloidal gold nanoparticles, etc.
HTML, REST API, Bulk download

ChEMBL

NM: Fullerenes , Metal oxides
HTML, REST API, Bulk download

ArrayExpress

Gene expression data
NM: carbon nanotubes, quantum dots, graphene oxide, zinc oxide, silver and gold nanoparticles.
HTML, REST API, Bulk download

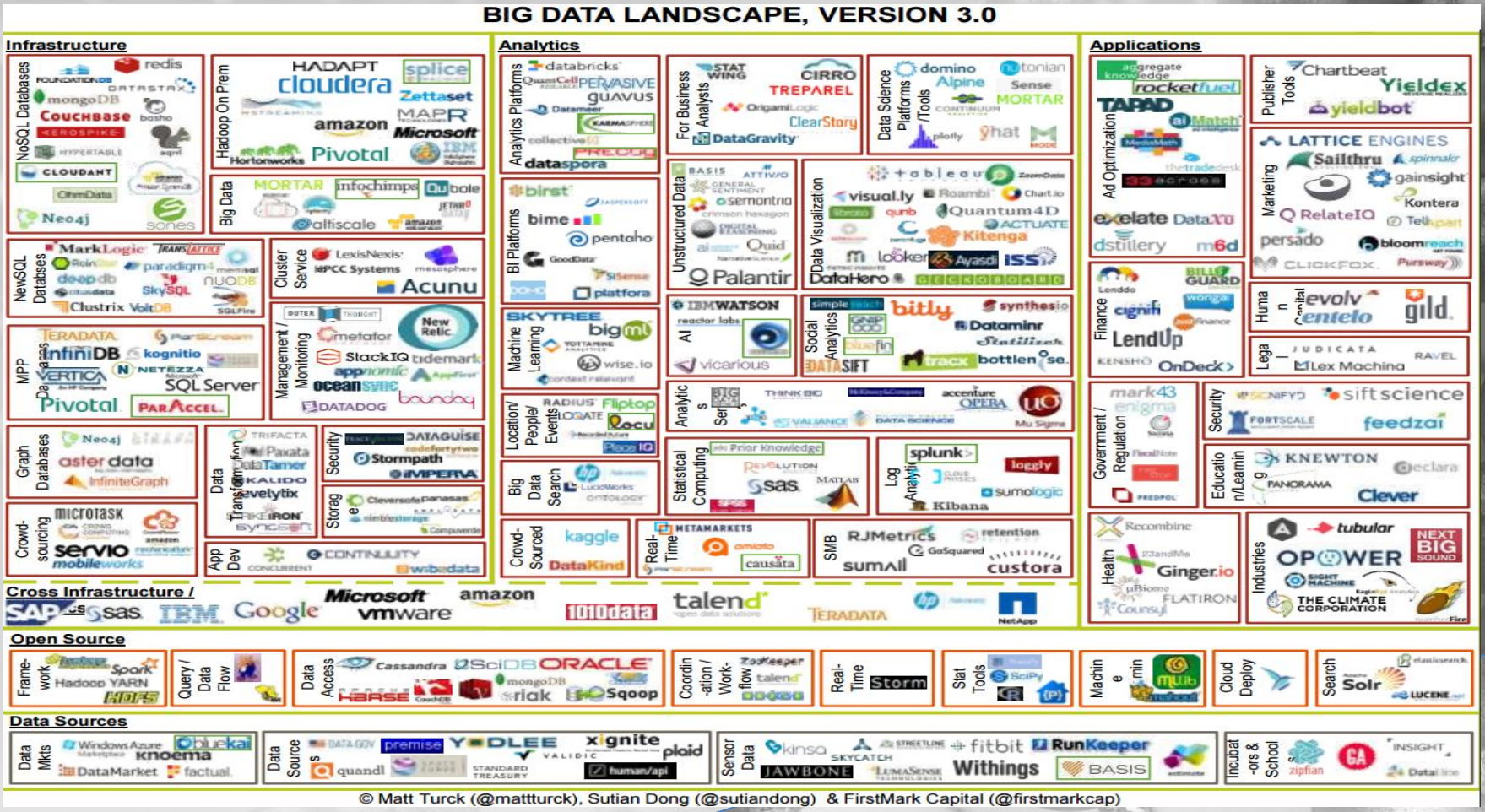
ctd™

Comparative Toxicogenomics Database
Includes nanomaterial related data. ***HTML***

ECHA EUROPEAN CHEMICALS AGENCY

The ECHA Dissemination site. Registered chemical substances under REACH, including NM. ***HTML only***

Data management and analytics



Data integration options

- Databases A & B
- Match entities
  - Common data model
  - Distinct data models, links
  - Distinct data models, joins
- Retrieve query results
  - Index time merging
  - Query time merging
  - Hybrid - Where practical, content is indexed centrally. Repositories for which this is not cost effective (or simply not possible) are federated to at query time.

	Single storage	Separate storage
Common data model	*Typical data warehouse*	*WWW*
Distinct data models	*Data lake*	*Toxygates * (2013) EMBL-EBI RDF ** Platform (2013)*

* http://toxygates.nibio.go.jp/

** http://www.ebi.ac.uk/rdf/

A **data lake** is a large storage repository and processing engine. They provide "massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs"

eNanoMapper

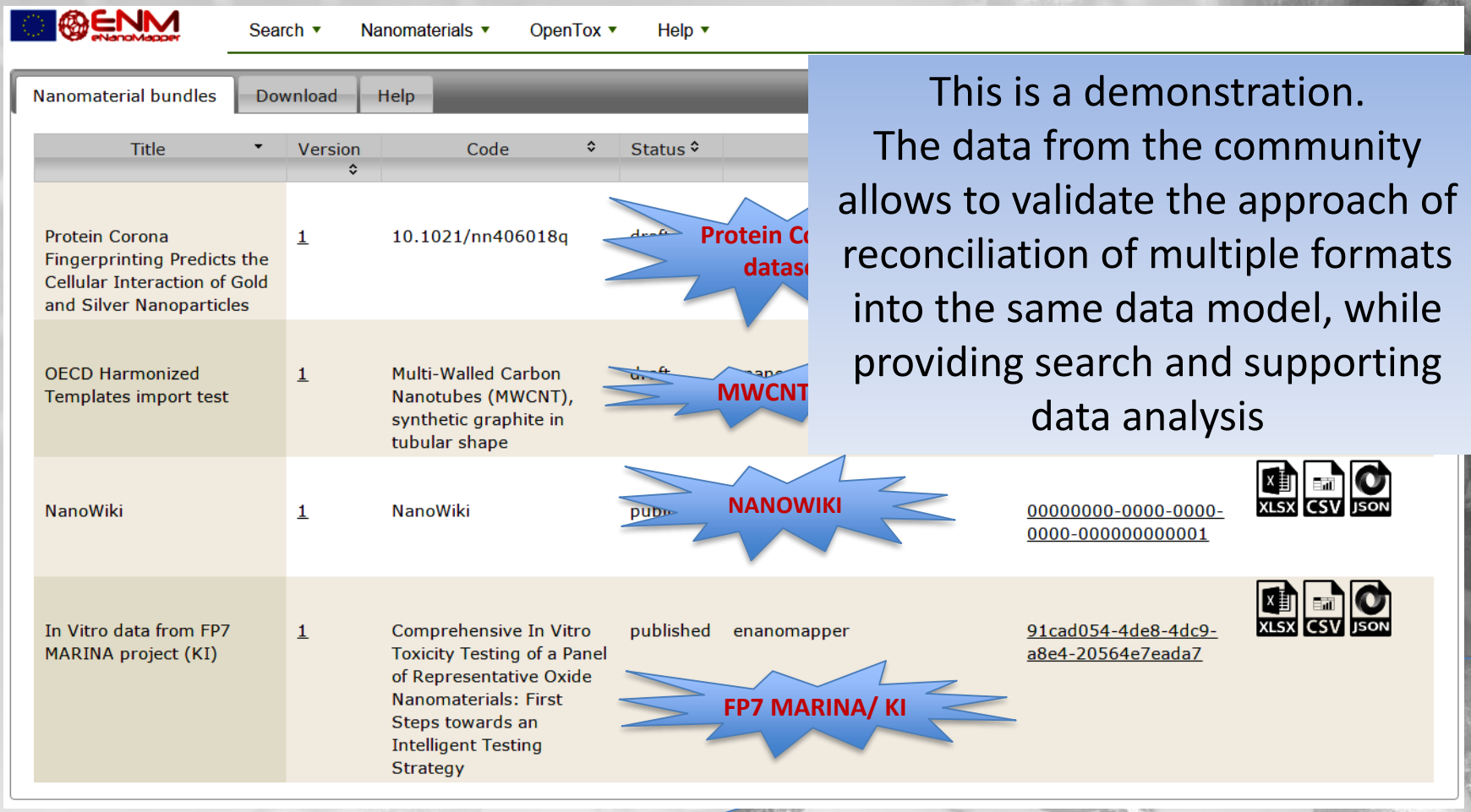
Identifiers for materials and measurements

- No unique identifiers for NM
  - Note *"Substance Identification"* is also an issue for chemicals in REACH
- However
  - NM are composed of chemical structures (hence identifiers available)
  - Crystallography , material structure
  - NM specific categorization exist (e.g. NanoCarbon nomenclature)
  - Measurements, protocols, endpoints, instruments, assay conditions – annotate by ontology entries

eNM eNanoMapper

TOOLS AVAILABLE

data.enanomapper.net



This is a demonstration.
The data from the community allows to validate the approach of reconciliation of multiple formats into the same data model, while providing search and supporting data analysis

eNanoMapper data model

- Flexible
  - General structure to describe measurements, but NO fixed fields for endpoints, experiment conditions, etc.
- Fields annotated by ontology entries
- Allows conversion between different formats and data models
  - Necessary for data integration!
- Programmatic access (API)
  - i.e. support for data analysis
- User friendly interface
  - Consuming the API (e.g. JavaScript)

eNM data model = generic description of any measurement. Does not specify what to record to describe particular experiment.

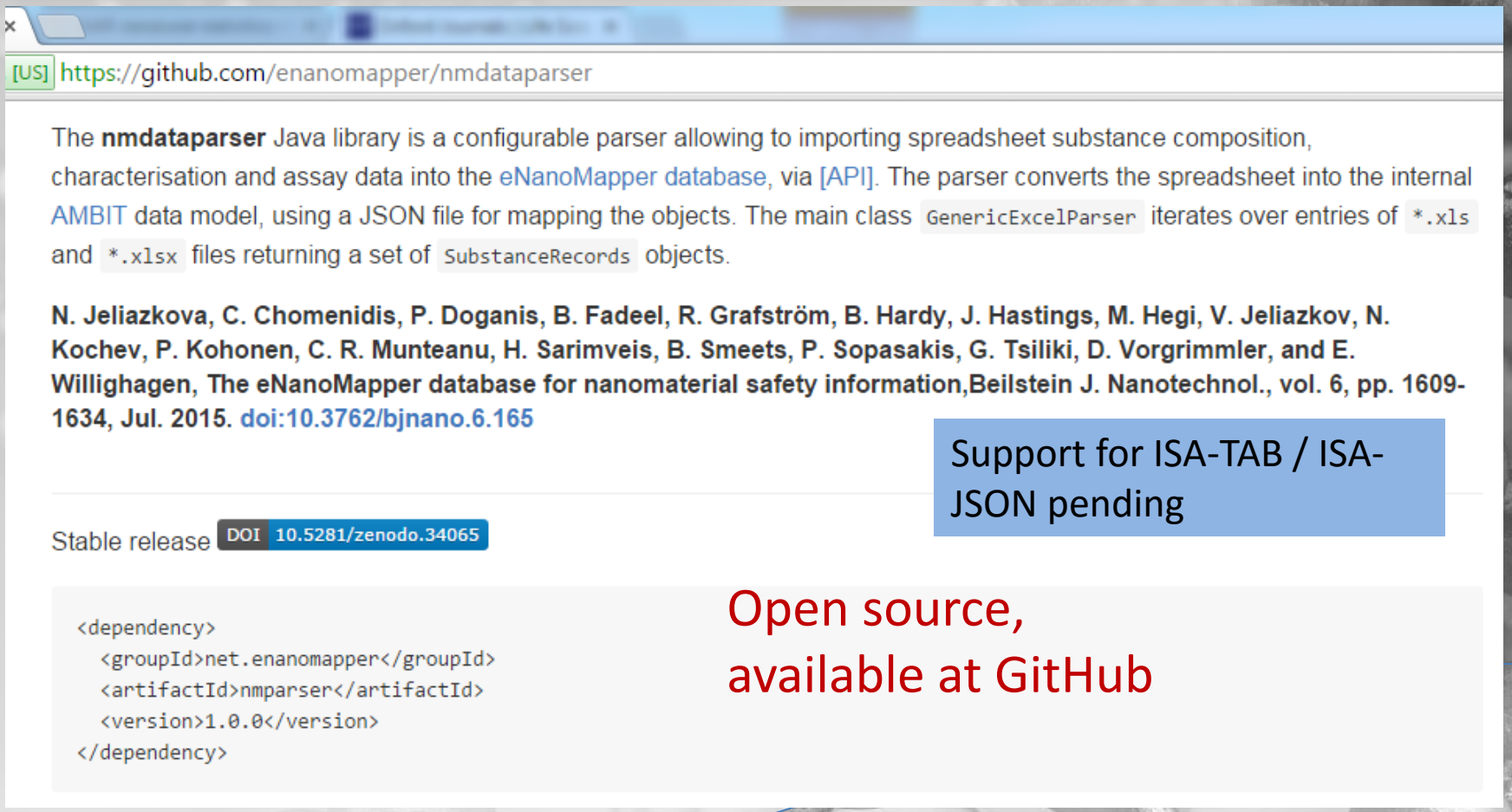
- The later is a domain specific scientific question
- Related work: CoDATA UDS , zeta potential pilot by US NanoWG, OECD WPMN, the NanoSafety cluster templates

How to represent the selected "aspects of the reality" in a database or ontology is a data representation question

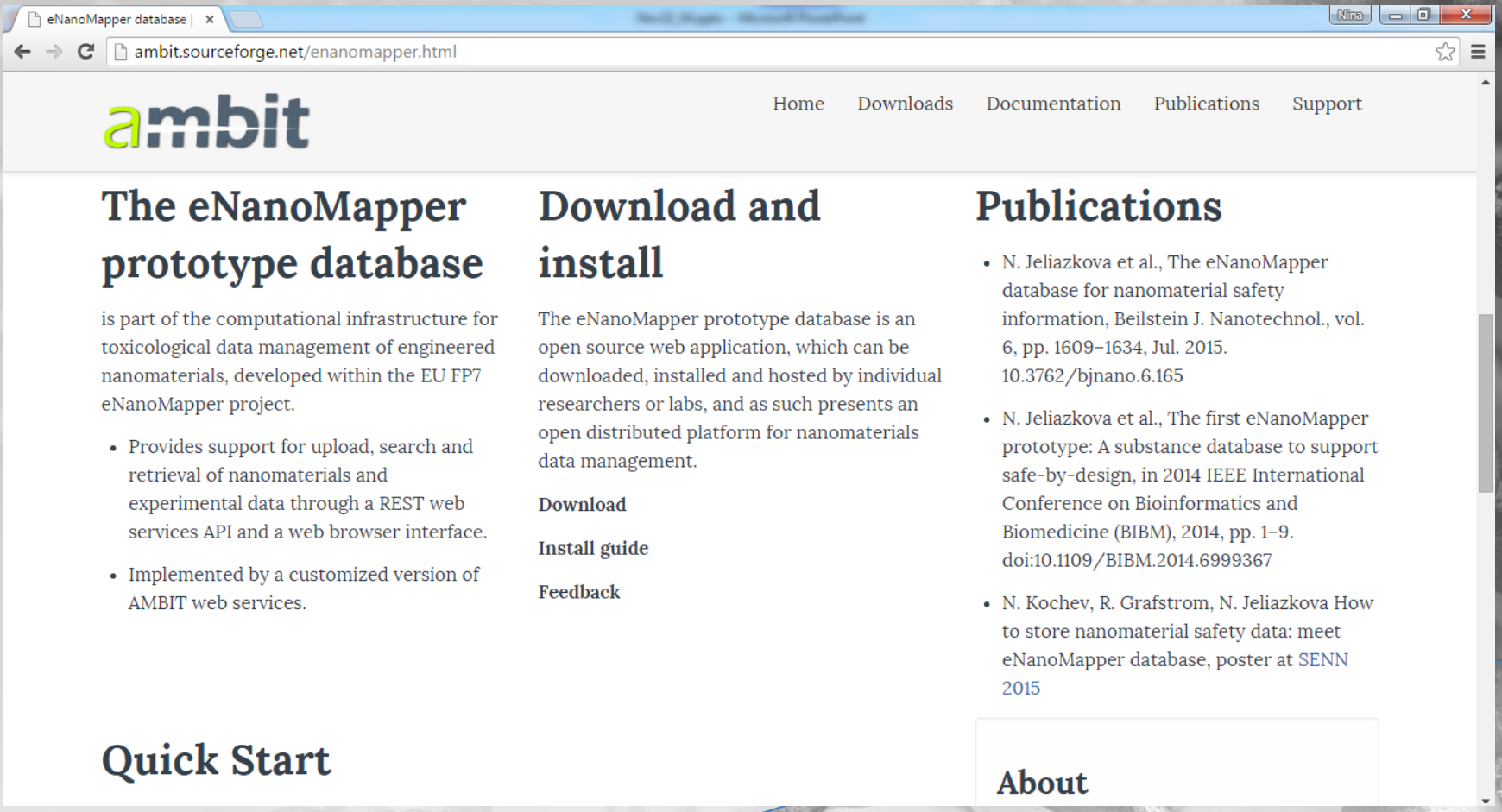
- Computer science, logic, data modelling; informed by the expected usage

eNM eNanoMapper

Format conversions, ISA schema



Downloadable



Published

- N. Jeliazkova, et al. **"The eNanoMapper database for nanomaterial safety information,"** *Beilstein J. Nanotechnol., vol. 6, pp. 1609–1634, Jul.* ***2015****.*

- N. Jeliazkova, et al. **"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.*

More details on eNanoMapper?

For any additional information, please ask us or download our publicly available Deliverables.

Douglas Connect Working communities

in silico toxicology

IOFA consult

Karolinska Institutet

VTT

EMBL-EBI

Maastricht University

misvik biology

eNanoMapper (Grant Agreement no. 604134) is a project supported by the European Commission through the Seventh Framework Programme (FP7).

eNM eNanoMapper

Questions?

THANK YOU!

eNM
eNanoMapper