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)



2 February 2016

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;

<image/> <section-header><section-header><section-header><section-header><section-header><complex-block><section-header><complex-block></complex-block></section-header></complex-block></section-header></section-header></section-header></section-header></section-header>		 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. 	
2 February 2016	***	ENN	1

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 id

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**Application**ki**programming** protocols; risk assessment; grouping, safety-bydesign **interface (API)**

Different views of the data

Requires "preadsheet" or matrix view of data. The experimental data in the public datasets is usually not in a for **substance dossier** fergin multiple values, conditions, similar experiments into matrix form is a challenge).



Data sources 1. Excel spreadsheet examples

1) Data transfer between FP7 NANoREG and eNanoMapper (WP3) Karolinska Institutet , Prof. Bengt Fadeel lab (NANoREG WP5)

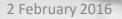
- cytotoxicity assessment of the entire panel of NANoREG nanomaterials
- All 19 nanomaterials were obtained from the JRC nanomaterial repository

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

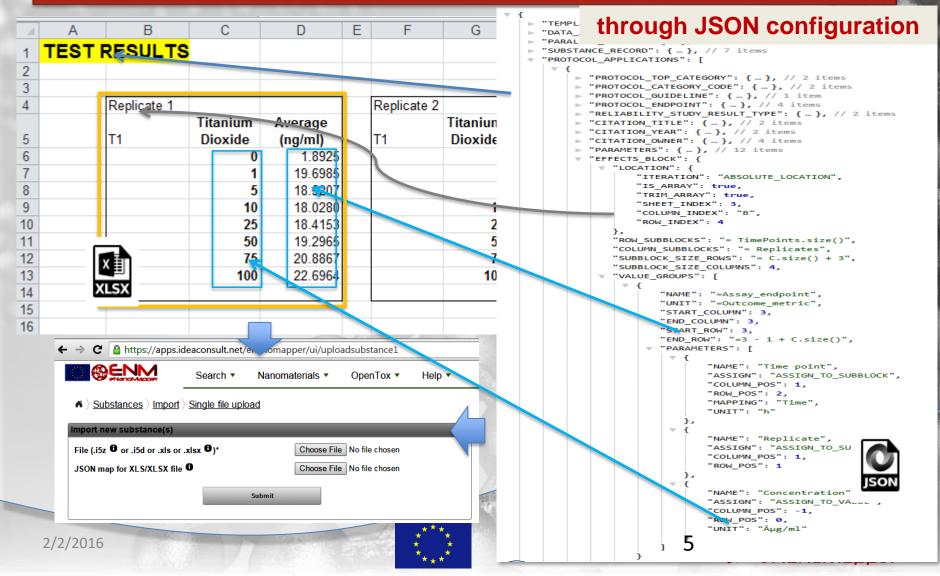
Karolinska Institutet, Prof. Bengt Fadeel lab

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 CONDITIONS	Please complete the details below as far as pos													
In-Vitro Template	While we need to to standardise MARINA toxicology data reco	ords as far as possible for compatibili	ty, som											
	You can add additional items below where necessary for furt In the notes area or adjacent to data tables, add annotations y			1.										
TEST and END POINT - GENERAL INFO	in the notes area of aspectitic data tables, and antonionis i													CONTRACT OF
MARINA Work Package:	WP09													
MARINA Partner ID: Test facility - Lab name etc:	P32 - Ki Division of Molecular Toxicology													
Work conducted by:	email address:													
Test / Assay End-Point short description: Enter full description in the covering TMDF - assay description form,	ELISA / TNF-a release in culture medium					-						-	-	
		A B	C	D	E F	G	н	J	K	L	M N	0	P	Q
End-Point Outcome metric (ie % viability, %cell death etc): (indicate how EP is derived)	TNF-a (ngimi)	TEST RESULTS												
		ILSI KLSULIS												and the second se
SOP - Protocol Name - ID (see project protocol ID list): (or add path/link to protocol on MARINA server)	M3Documents_per WPWP9/Task 9.1WANOMMUNE_OHB_FINAL_2011	pd												
Test start date (dd/mm/yyyy):	11/1/2012													
Test end date (dd/mm/yyyy):	4/30/2013	D. F. J. J.			0 5 5 6			D			D. F.	-		
TEST SUBSTANCE		Replicate 1			Replicate 2			Replicate			Replicat			
	Titanium Dioxide TIO2		Titanium	Average		Titanium	Average		Titanium	Average		Titanium	Average	
CAS No:	Ttanium Doxide 1102	T1	Dioxide	(na/ml)	T1	Dioxide	(na/ml)	T1	Dioxide	(na/ml)	T1	Dioxide	(ng/ml)	A statistical statistics of the statistical statistics
Standard MARINA Nanomaterial Code & Name:	NM-103 (TiO2) NM-103 NPO_\$495		DIOXIGE			DIOXIGE			Dioxide			Dioxide		A CONTRACT OF A
(See MARINA Materials list) Highest concentration, inc units:	100 µo/mi	-	0	1.8925		0	-0.3425		0	2.189		0	9.502	
DISPERSION			1	19.6985		1	5.9887		1	2.474		1	8.808	
			5	18.5207		5	5,7696		5	2.002		5	8.957	
Specify the standard dispersion protocol used: (or otherwise specify the dispersion technique used)	NANOGENOTOX Dispersion protocol (http://www.nanogenotox.eu/files	PC	5			10		-			-			
			10	18.0280		10	5.3011		10	1.786		10	8.306	and the second
	water with 0,05% RRA		25	18.4153		25	5.1602		25	1.570		25	8.244	
Aids used to disperse - Y / N:	Sonication: Stirring: Y		50	19.2965		50	5.2292		50	1.735		50	9.507	
Treatment concentration series (C) (µg/ml):	C C4 C5					75			75					
Alter or add as necessary	10 25		75							1.290		75		
CELL LINE/TYPE	X =		100	22.6964		100	7.2811		100	1.206		100	15.438	and the second se
Short-Name:	HMDM (primary ce													The second s
Full specific name (note any line variants or related IDs): Supplier:	ATCC													
CELL CULTURE CONDITIONS														
	Y SY									ar 005				and the second se
Medium (Supplier/Lot No.): Serum (Supplier/Lot No.):	PBS (Gbco/41310									20 Vill				ARCHINESE - 1 - 1 - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 7
Serum concentration in growth medium: Serum concentration in treatment medium:	10									e				
TIMELINE										1 N			5	
Time points (hours - or specify any other units):	71					-	and the second			1 A A A A A A A A A A A A A A A A A A A				and the second
Alter or add as necessary	24					1002	2.57				and the second s		100	
NOTES - including any deviations from SOP; other observation	s, variations etc. Add any information that will assist in the use a	ind interpretation of the data					128					35.7	1.10	
F H Test conditions Raw data Test results Te							1000						Contraction of the second	



Mapping the spreadsheet content into the data model



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

```
TEMPLATE_INFO": { ... }, // 3 items
"DATA_ACCESS": { ... }, // 9 items
"DATA_ACCESS": { ... }, // 9 items
"PARALLEL_SHEETS": [ ... ], // 1 item
"SUBSTANCE_RECORD": { ... }, // 7 items
"PROTOCOL_APPLICATIONS": [
          PROTOCOL_TOP_CATEGORY": { ... }, // 2 items
         "PROTOCOL_CATEGORY_CODE": { ... }, // 2 items
         "PROTOCOL GUIDELINE": { ... }, // 1 item
          'PROTOCOL ENDPOINT": { ... }, // 4 items
          RELIABILITY_STUDY_RESULT_TYPE": { ... }, // 2 items
          CITATION_TITLE": { ... }, // 2 items
          CITATION_YEAR": { ... }, // 2 items
         "CITATION_OWNER": { ... }, // 4 items
          'PARAMETERS": { ... }, // 12 items
           EFFECTS_BLOCK": {
               LOCATION": {
                  "ITERATION": "ABSOLUTE LOCATION",
                  "IS_ARRAY": true,
                  "TRIM ARRAY": true,
                  "SHEET_INDEX": 3,
                  "COLUMN_INDEX": "B",
                  "ROW_INDEX": 4
              "ROW_SUBBLOCKS": "= TimePoints.size()",
              "COLUMN_SUBBLOCKS": "= Replicates",
              "SUBBLOCK_SIZE_ROWS": "= C.size() + 3",
              "SUBBLOCK_SIZE_COLUMNS": 4,
               VALUE_GROUPS": [
                       "NAME": "=Assay_endpoint",
                       "UNIT": "=Outcome_metric",
                       "START_COLUMN": 3,
                       "END COLUMN": 3.
                       "START_ROW": 3,
                       "END ROW": "=3 - 1 + C.size()",
                       'PARAMETERS": [
                                "NAME": "Time point",
                                "ASSIGN": "ASSIGN_TO_SUBBLOCK",
                                "COLUMN_POS": 1,
                                "ROW_POS": 2,
                                "MAPPING": "Time",
                                "UNIT": "h"
                           ъ.
                                "NAME": "Replicate",
                                "ASSIGN": "ASSIGN_TO_SUBBLOCK",
                                "COLUMN_POS": 1,
                                "ROW_POS": 1
                                'NAME": "Concentration",
                                "ASSIGN": "ASSIGN_TO_VALUE",
                                "COLUMN POS": -1,
                                "ROW_POS": Ø,
"UNIT": "Âμg/ml"
```

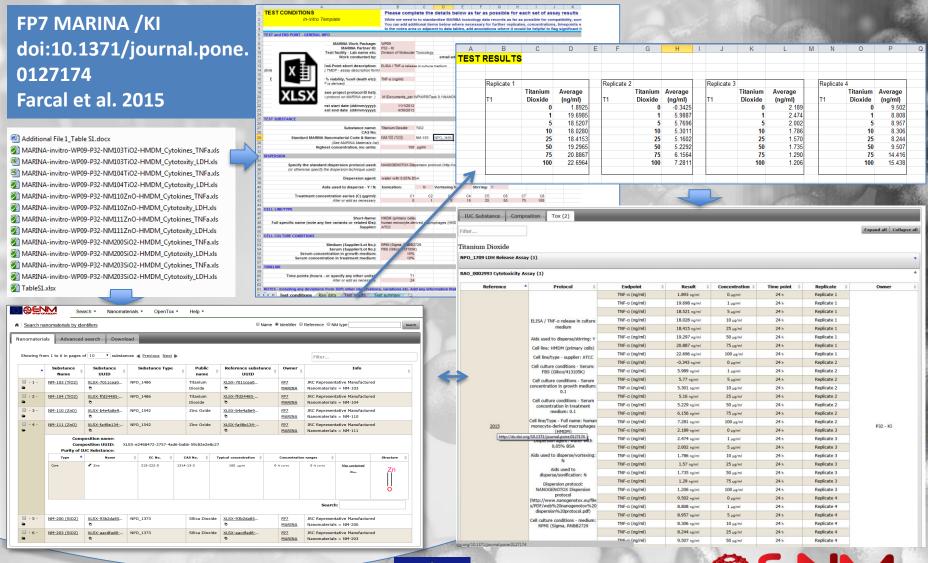


Data imported (MARINA/KI) **GENM** Search • Nanomaterials • OpenTox • Help • ○ Name ● Identifier ● Reference ● NM type Search nanomaterials by identifiers Search Nanomaterials Advanced search Download Showing from 1 to 6 in pages of 10 substances <u>Previous</u> <u>Next</u> Filter... Substance Substance Public Reference substance Info Substance Type Owner Name UUID name UUID -1-NM-103 (TiO2) XLSX-7011cea0... Titanium XLSX-7011cea0... <u>FP7</u> JRC Representative Manufactured NPO_1486 • Dioxide • MARINA Nanomaterials = NM-103 - 2 -NM-104 (TiO2) XLSX-ffd24485-... NPO_1486 Titanium XLSX-ffd24485-... <u>FP7</u> JRC Representative Manufactured • C Dioxide MARINA Nanomaterials = NM-104 - 3 -NM-110 (ZnO) XLSX-b4e4a8e9... NPO 1542 Zinc Oxide XLSX-b4e4a8e9... FP7 JRC Representative Manufactured • • MARINA Nanomaterials = NM-110 - 4 -NPO_1542 JRC Representative Manufactured NM-111 (ZnO) XLSX-fa48e134-... Zinc Oxide XLSX-fa48e134-... FP7 **P** ۲ • MARINA Nanomaterials = NM-111 **Composition name: Composition UUID:** XLSX-e2468472-2757-4ad6-babb-59c82e2e8c27 Purity of IUC Substance: Туре Name EC No. CAS No. Typical concentration Concentration ranges 4 Structure ₿ Zno 215-222-5 1314-13-2 Core 100 µg/ml 0 % (w/w) 0 % (w/w) Also contained Zn in... O Search: - 5 -XLSX-93b2da85... NM-200 (SiO2) NPO_1373 Silica Dioxide XLSX-93b2da85... <u>FP7</u> JRC Representative Manufactured P • • Nanomaterials = NM-200 MARINA - 6 -XLSX-aacdfad8-... NPO_1373 Silica Dioxide XLSX-aacdfad8-... JRC Representative Manufactured NM-203 (SiO2) FP7 P ۲ ۵ Nanomaterials = NM-203 MARINA





From spreadsheets to DB content







Data sources 2. PDF files

	Summary of physico-chem	nical data		Physico-chemical property Shape	Protocol Endpoint TEM Shape	Description of physico-chemical properties Units Experimental conditio Experimental condition (Experimental Condition) Primary subunits tend to be more or less equi-axed. The 3D structure suggests they are spherical or ellipsoidal
Physico-chemical	Description of physico-chemical	Source of information	Date	3 Shape	TEM Shape	Primary solutions territor to be indee or insiste our actor. The solution is solution solution is solution to the provide a more fractal-like structure [1-2] and minor amounts of singlet spheroidal particl 10±57
property (if applicable)	properties (if applicable)		20.01.14	4 Core size distribution 5 Core size distribution 6 Core size distribution	TEM Size TEM Size XRD Size	20- 36.3 V {
Shape	TEM: Primary subunits tend to be more or less equi-axed. The 3D	[1] NANOGENOTOX: NRCWE data	28.01.14	7 Core size distribution	TEM size <100 nm	■ "TEMPLATE_INFO": { }, // 3 item
~	structure suggests they are	[2] JRC technical report		8 Core size distribution	TEM Size <50 nm	<pre>v "DATA_ACCESS": {</pre>
	spherical or ellipsoidal [2]	[L] me recimical report		9 Core size distribution	and	"ITERATION": "ROW SINGLE",
~	Aggregates/agglomerates have a			10 Core size distribution 11 Aspect ratio	TEM Aggregates aggion TEM Aspect ratio	1.3 "SHEET INDEX": 1,
-	more fractal-like structure [1-2] and			12 Surface morphology/topography 13 Specific surface area 14 Specific surface area	BET Specific surface and BET Specific surface and	"START ROW" - 5
DE	minor amounts of singlet			15 Crystalite structure 16 Core material (elemental composition)	XRD EDS Ti	"END_ROW": 10,
	spheroidal particles [1]			17 Core material (elemental composition) 18 Overall material composition (including degree of		"START_HEADER_ROW": 2,
Core size distribution	TFM: 110 + 57 nm [1]: from 20 to	[1] NANOGENOTOX-	28.01.14	20 Overall material composition (including degree of		"END_HEADER_ROW": 3,
hus-cham ch	naracterisation of re	oprocontativo ma	atorials	21 Overall material composition (including degree of 22 Overall material composition (including degree of 23 Overall material composition (including degree of	purity, IEDS AI	"ALLOW_EMPTY": true,
ilys-chemici		epiesentative ma		24 Overall material composition (including degree of 24 Overall material composition (including degree of		RECOGNITION": "BY_INDEX"
RC Nanomat	erials Repository			25 Overall material composition (including degree of 26 Overall material composition (including degree of		···· },
				27 Overall material composition (including degree of		xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
uan Riego-Si	ntes			29 Coating 30 Functional groups present on the coating	TGA	org ▼ "PROTOCOL_APPLICATIONS": [
				31 Zeta potential 32 Hydrodynamic radius 33 Hydrodynamic radius	DLS ZETA POTENTIAL DLS HYDRODYNAMIC R DLS HYDRODYNAMIC R	
ttps://ec.eu	<u>ropa.eu/jrc/en/scie</u>	ntific-tool/jrc-		33 Hydrodynamic radius 34 Surface area 35 Hoavy metals contaminants	BET Nitrogen BET adsor ICP analysis P	wptic TITLE": {
		, , <u>, ,</u>				"COLUMN_INDEX": "Q"
<u>anomaterial</u>	<u>s-repository</u>					},
						▼ "PROTOCOL_GUIDELINE": {
	BET: 10.03 m2/g (room temp) [2]	Deliverable 4.4				▼ "guideline1": {
		[2] JRC technical repor ← → C	https://apps.ideaconsu	.ult.net/enanomapper/ui/uploadsubsta	ncel	"COLUMN_INDEX": "
						},
			Search Search	ch 🔹 Nanomaterials 🔹 OpenT	Fox ▼ Help ▼	<pre>"PROTOCOL ENDPOINT": {</pre>
Search • Nanomatorials • ()	ipenTox * Help *	and the second s				"COLUMN INDEX": "P"
anomaterials by identifiers) NWRI-3d8aaa66-97d9-35	,	₼) Su	ubstances > Import > Single :	file upload		
Intraterials by Methiles / Hittle Subalactory 129:55	TREAT THE ANALYSIST / JUNKE & Maine & Evening Resident		and and a minister of angle	<u></u>		T "PARAMETERS": {
nce Composition P-Chem (4)		Import	new substance(s)			▼ "Dispersion": {
	Expand a	a Collapse all	new substante(s)			"COLUMN_INDEX": '
e distribution (Granulometry) (2)		File (.i5	iz 🖯 or .i5d or .xls or .xlsx 🖯)*	Choose File No file cl	hosen	},
Reference A Proto	col 0 Endpoint 0 Result 0 Ow	ner e	·		. 🛛 🗋	▼ "Sample Provider": {
<u>D01</u>	PARTICLE SIZE [42, 90] nm Nam	JSON m	nap for XLS/XLSX file	Choose File No file cl	nosen	"COLUMN_INDEX": "
D01	PARTICLE SIZE 267 mm Name	wiki	_		-	- · · · · · · · · · · · · · · · · · · ·
dy(s) (1 to 2)	4 Pre	iaus Next 🕨		Submit		3.
erial aspect ratio/shape (1)		•				► "EFFECTS": [] //
erial specific surface area (1)						}
Reference A Proto	col 0 Endpoint 0 Result 0 Ow	ner 0				JS JS
001	SPECIFIC_SURFACE_AREA 10 == 2/9 Name	wiki	100	18-PM		}
dy(s) (1 to 1)	4 Pre:	ious Next 🕨	and the second sec	M		Contraction of the second seco
	111					
		2.4.7	★ ^ ★ D			
0. F. k	1.0	and the second	* *			
2 February 20	16		* *			

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

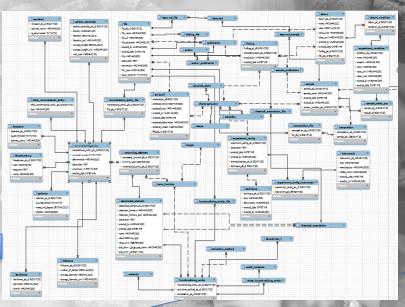
- MySQL dump provided to eNanoMapper
- Domain model class names and attributes <u>https://cdebrowser.nci.nih.gov/CDEBrowser/</u>
- Definitions for caNanoLab concepts are maintained in the NCI Enterprise Vocabulary Services <u>https://ncit.nci.nih.gov/ncitbrowser</u>
- Code: <u>https://github.com/NCIP/cananolab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

rch substances by endpoint data) Hit list Hit list CNLB-fe79e508-15695885 McCarthyNL2005-0 CNLB-816a4cbc 15695885 m 4.5. Particle size distribute a 4.8. Water solubility (S) [14] NLB-91f65398-CNLB-09919cc2 NEU-LWangJNBT2008-0 CNLB-681de79f-15695950 CNLB-88579c37 15695943 AVillanuevaNT2009-0 AVillanuevaNT2009-0 UAM CSIC IMDEA-CNLB-88cf623c-UAM CSIC IMDEA 15695943 NIOSH-VWalkerTAP2009 CNLB-007b04f6-NTOSH-VWalkerTAP2009-02 KI-HKarlssonCRT2008-01 CNLB-84e50771-1481113 nCRT2008-01 KI-HKarlssonCRT2008-0 CNLB-1d676b8b 1481113 CNLB-594db8b0 1481113

Mapping into eNanoMapper data model

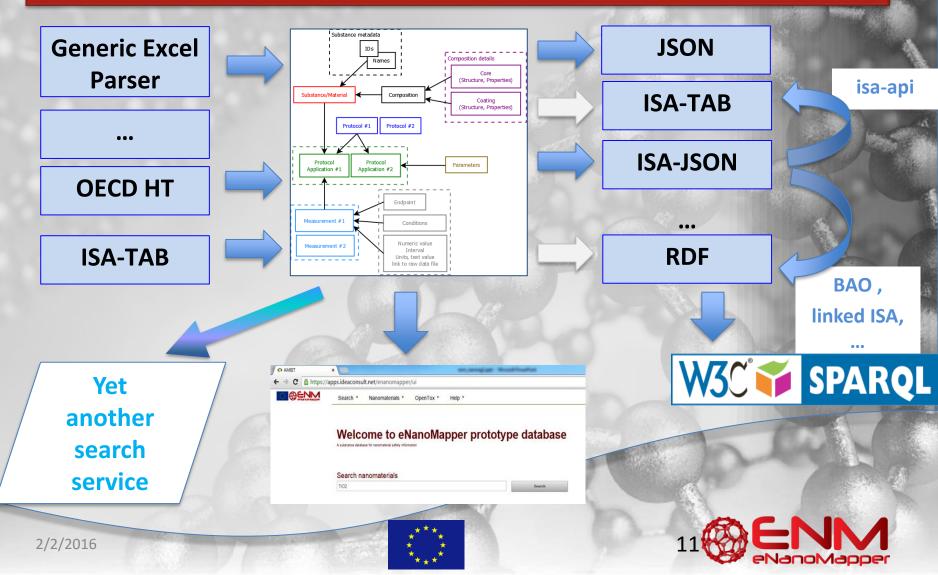




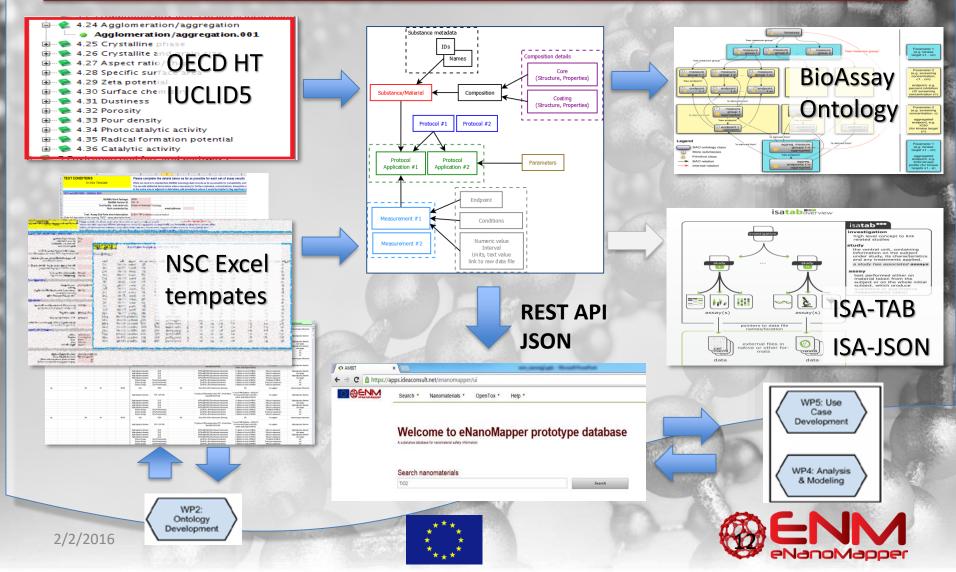




eNanoMapper data format conversion



eNanoMapper supported data formats



isatab[®]

investigation

study

assav

related studies

high level concept to link

the central unit, containing

information on the subject

and any treatments applied.

test performed either on

material taken from the

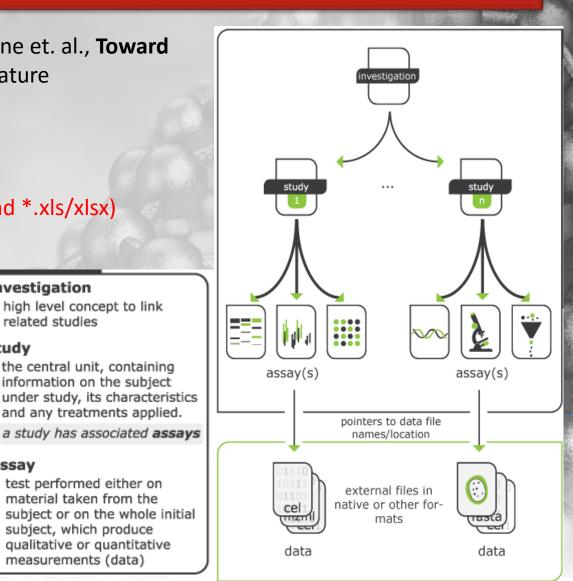
subject, which produce

measurements (data)

qualitative or quantitative

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)



2 February 2016

isatab (Investigation file)

	А	в
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 I
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

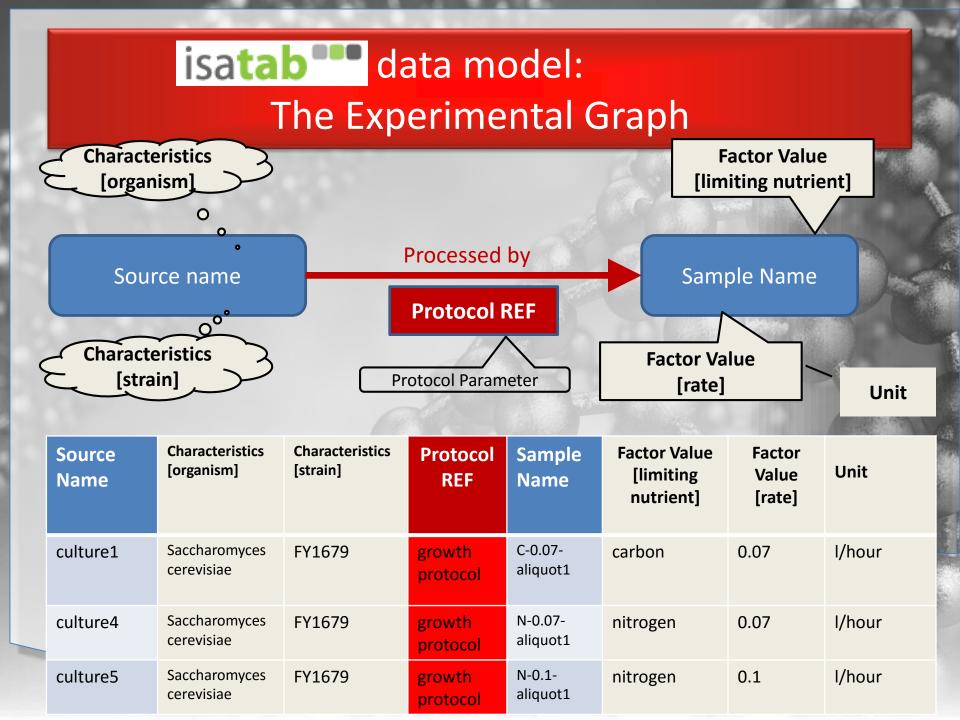
Study Person Roles Term Source REF		1
Comment[Study Person REF]		
STUDY		
Study Identifier	BII-S-2	a land
Study Title	A time course analysis of transcription response in yeast treated with r	hand the
Study Description	Comprehensive high-throughput analyses at the levels of mRNAs, prot	2 DAY
Comment[Study Grant Number]		
Comment[Study Funding Agency]	pointer to the study fil	
Study Submission Date	4/30/2007 pointer to the study in	e
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	pointers to the as	sav files

Each study contains one or more assays

			And the second sec
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_0	0 http://purl.obolibrary.org/o	42







ISA-TAB-Nano

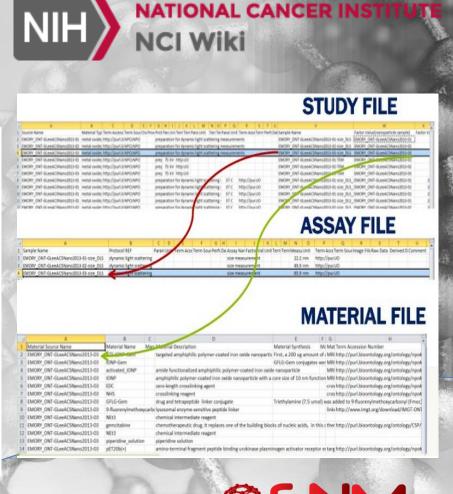
The ISA-TAB-Nano project is an effort of the National Cancer Institute (NCI) National Cancer Informatics Program (NCIP) Nanotechnology Informatics Working Group (Nano WG)

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







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	М	A;B;C	
M-lab-1	Μ	A;B	covalent linkage
M-lab-1	Μ	A; C	covalent linkage
M-lab-1	А		
M-lab-1	В		
M-lab-1	С	Material N	M is comprised of A, B, and

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



ISA-JSON project isatab 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



2 February 2016

ISA-JSON schemas



https://github.com/ISA-tools/isaapi/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 International schema ison

```
"$schema": "http://json-schema.org/draft-04/schema",
"title" : "ISA investigation schema",
"description" : "JSON-schema representing an investigation in the ISA model",
"type" : "object",
"properties" : {
    "identifier" : { "type" : "string" },
    "title" : { "type" : "string"},
    "description" : { "type" : "string"},
    "submissionDate" : { "type" : "string", "format" : "date-time"},
    "publicReleaseDate" : { "type" : "string", "format" : "date-time"},
    "commentCreatedWithConfiguration" : {
        "$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

This repository Search	Pull requests Issues Gist		+- 🖬-
enanomapper / isa-api forked from ISA-tools/isa-api		O Unwatch → 18 ★ Star	0 % Fork 2
↔ Code 🕅 Pull requests 0 🗉 Wiki ≁ Pulse	III Graphs 🔅 Settings		
Branch: master - isa-api / isatools / schemas / isa	_model_version_1_0_scher	mas / material / New file Fi	nd file History
This branch is 8 commits ahead, 149 commits behind ISA-tools:m	naster.	🕄 Pull requ	uest 主 Compare
ntk73 added constituent_schema.json		Latest commit 951f71	7 on Dec 16, 2015
 constituent_schema.json added constituent_schema.jso	on		a month ago
material_schema.json Added basic properties of ma	terial json-schema		a month ago
2016 GitHub, Inc. Terms Privacy set y Contact Help	0	Status API Training Shop I	Blog About Pricing
Contributing new	IOSL	N schema correspor	nding
extension to isa-api		to ISA-TAB-Nano	
(under development		material file	
by eNM team)			
		and the second se	
Prebruary 2016	****		

eNanoMapper: ISA (v.1) Java classes

- 🖻 🚺 Assay.java
- 👂 🚺 Comment.java
- > 🚺 Component.java
- 🖻 🚺 Data.java
- 🖻 🚺 Factor.java
- FactorValue.java
- Investigation.java
- Material.java
- MaterialAttribute.java
- > D MeasurementType.java
- I OntologyAnnotation.java
- D OntologySourceReference.java
- > D Organization.java
- > 🚺 Parameter.java
- > 🚺 Person.java
- > D Process.java
- > 🚺 Protocol.java
- > D Publication.java
- 🛛 🚺 Role.java
- 🛛 🚺 Sample.java
- Source.java
- 🗴 🗕 Study.java
- J TechnologyType.java

2 February 2016

> J) Value.java

Investigation.java

- Investigation
 - commentCreatedWithConfiguration
 - commentLastOpenedWithConfiguration
 - description
 - identifier
 - ontologySourceReferences
 - o 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<OntologySourceReferences ontologySourceReferences = new A
@JsonProperty("publications")
public List<Publication> publications = new ArrayList<Publication>();
@JsonProperty("studies")
public List<Study> studies = new ArrayList<Study>();

🔻 🚺 Study.java

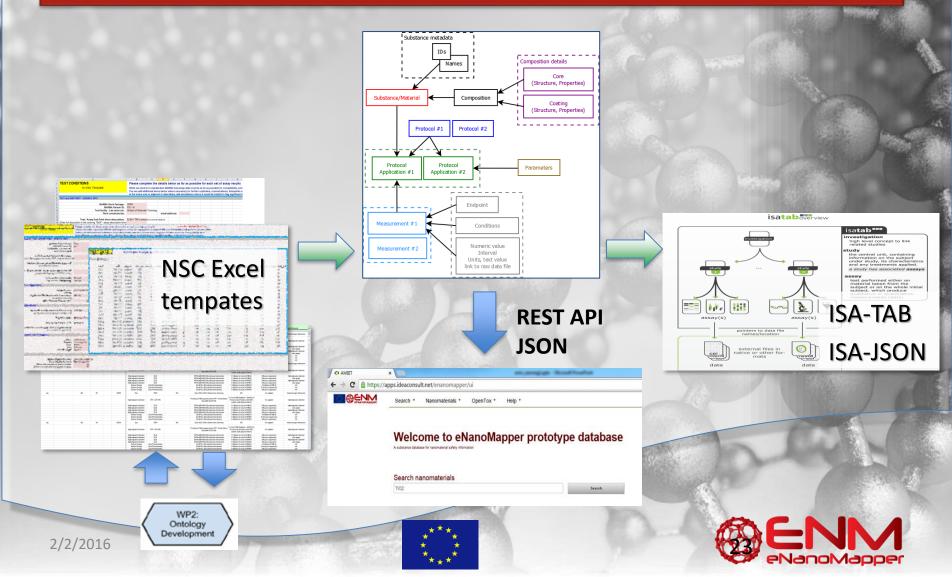
Study

- assays
- description
- identifier
- o people
- o 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<Ontology @JsonProperty("protocols") public list Protocols protocols - new Arraylist Protocols():

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



Searchability



Search • Nanomaterials • OpenTox •

Welcome to eNanoMapper prototype database

Help •

A substance database for nanomaterial safety information

Search nanomaterials

TiO2

Search

Search, browse, upload

Search nanomaterials by identifier O Search nanomaterials by citation O	Search nanomaterials by physchem parameters or biological effects	Search nanomaterials
Free text search	Browse nanomaterials and studies O Data exchange via <u>REST</u> <u>API</u> C ^A O	Data import Supported import formats: OECD HT , Excel spreadsheets

Help: Nanomaterials
The nanomaterials
are considered a special case of
substances
. See <u>doi:10.3762/bjnano.6.165</u>







Search by physchem and bio effects

GENM

Search < Nanomaterials <

OpenTox • Help •

▲ > Search substances by endpoint data > Hit list

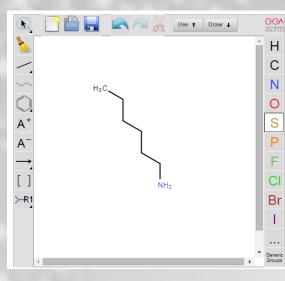
	e results	Hit list	Download	Н	elp							
P-C	Them				- 1							
	🖀 4.1. Appearance (<u>S</u>) [1]	Showing	from 1 to 10 in	pages o	of 10 🔹 substan	ces 📹 <u>Previous</u>	<u>Next</u> 🕨		Filter	Filter		
	4.26. Nanomaterial crystallite and grain size (<u>S</u>) [105]		Substance N	ame	Substance	Substance	Public	Reference	Owner	Info		
	■ 4.27. Nanomaterial aspect ratio/shape (<u>S</u>) [9]	^		*	UUID 🔶	Туре 🍦	name 🍦	substance 🍦 UUID	*			
	4.28. Nanomaterial specific surface area (<u>S</u>) [35]		Limbach2005	NM1	<u>NWKI-7998492c</u>	ENM_9000006	CeO2 I	<u>NWKI-7998492c</u>	<u>NanoWiki</u>	Composition = CeO2		
)	4.29. Nanomaterial zeta potential (<u>S</u>) [249]	- 1 -			6			6		DATASET = NanoWiki Has_Identifier = 161		
)	4.30. Nanomaterial surface chemistry (<u>S</u>) [368]		Compositio	on name	.					SOURCE = Limbach20		
	4.31. Nanomaterial dustiness (<u>S</u>) [1]		Compositi		D: NWKI-79984920	-3902-3cc2-9f92	cdfae53cfc?	02				
	4.5. Particle size distribution (Granulometry) (<u>S</u>) [513]			bstance								
	point name Units		Туре 🔺	Nan	ne . 	CAS No.	Typical concentration	Concentration	ranges 🍦	Structure		
	er endpoint value		Core	₽ Ceo2		1306-38-3	0 % (w/w)	0 % (w/w) (Also ntained		
Ento >= 1												
- 1	▼ 50 <= ▼ 60											
= Tox	▼ 50 <= ▼ 60							503	rch:			





Log in

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.

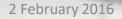
There is not yet a standardized approach for NM similarity, however a number of attempts for NM grouping and read across have been published recently.

howir	g from 1 to 3 in pag		🔹 entries 🔌	Previous Ne	<u>xt</u> ►	_		_	F	ïlter		Expo
	Diagram	CasRN	EC number	IUCLID 5 R	Names	Trade Name	UPAC nar	ne SMILES	🛊 Std. InChi key 🍦	Std. InChi 🛊	REACH registration dat	te Simila
1-				FCSV-97	Octadecylamine, octadecan-1- amine[REY.JJPSVUYRZGE UHFFFAOYSA- NIInCh1=IS/C18H39N/c1- 2-3-4-5-6-7-8-9-10-11-12- 13-14-15-16-17-18-19/h2- 19H2,1H3[stearylamine]1- octadecanamine	-	-	NCCCCC	REYJJPSVUYRZGE- UHFFFAOYSA-N	inChi=1S የ	•	0.72
		Identifiers	Substances	Datasets	_	_	_	_	_	-	_	-
	~~~~~	Showing f	rom 1 to 4 in	pages of 20	substances 4 President	vious Next 🕨				Filter		
			Substance Name		tance Substance T IID 🕴	ype Pub \$	lic name ¢	Reference substance 🍦 UUID	Со	atin	g +	Contained in as
2 -		- 1 -	S40.HDA	<u>FCSV</u> ବ	- <u>Oe</u> NPO_1892	540.		e <u>F</u>	rotein Corona Fingerprint redicts the Cellular Intera Gold and Silver Nanopartic	action of	ila. T = Cationic	coating
•		- 2 -	G60.HDA	FCSV ®	- <u>ac</u> NPO_401	G60.		е <u>в</u>	rotein Corona Fingerprint redicts the Cellular Intera Gold and Silver Nanopartic	action of	Classification = Cate ic	coating 🛛
		- 3 -	G15.HDA	<u>FCSV</u> ବ	<u>-c4f</u> NPO_401	G15.		е <u>в</u>	rotein Corona Fingerprint redicts the Cellular Intera Gold and Silver Nanopartio	action of	Classification = Cationic	coating
		- 4 -	G30.DDT@I	HDA FCSV	<u>-cb</u> NPO_401	G30.			rotein Corona Fingerprint redicts the Cellular Intera	_	Classification = Cationic	coating



### 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

Search * Nanomaterials * OpenTox • Demo • Help • Showing from 1 to 1 in pages of 10 substances 
Previous Next Filter... Public Reference Substance Substance UUID Substance Type Owner Info . Name name substance UUID . - 1 -FCSV-bc77c03d-FCSV-50cca421-... G15.AC nanoparticle G15.AC Protein Corona Fingerprinting Predicts the Classification = Anionic . ۰ • Cellular Interaction of Gold and Silver Nanoparticles.csv Composition name: Composition UUID: FCSV-bc77c03d-4e75-3fab-bb3d-17b983663819 Purity of IUC Substance: Typical concentration ۸ Name EC No. CAS No. Concentration ranges ÷ Structure Туре 4 4 & (2r)-2-Acetamido-3-Coating 0 % (w/w) 0 % (w/w) 0 % (w/w) Also contained Sulfanyl-Propanoic <u>in...</u> Acid.Pwkskimoespvia-Bypyzucnsa-N,Inchi=1s/C5h9no3s/C1-3(7)6-4(2-10)5(8)9/H4, 10h, 2h2, 1h3, (H,6,7)(H,8,9)/T4-/M0/S1. (2r)-2-Acetamido-3-Coating Sulfanylpropanoic Acid, (2r)-2-Acetamido-3-Mercapto-Propionic Acid, (2r)-2-Acetamido-3-Mercaptopropanoic Acid,N-Acetyl-L-Cysteine Core 8 [Au] 0 % (w/w) 0 % (w/w) 0 % (w/w) Also contained in... Au Core



VanoMapper

### Free text search supported by ontology annotated database

### Several technical options available

	and the set of the set of the					4		
	Search   Nanomaterials   OpenTox	Dem	no ▼ Help ▼	– Search for				
A > Free text search O All • Endpoint O Protocol O NM type size search							Scare	
Substances Advar	nced search Download		_	_	_	Endpoint	Protocol  NM ty	pe metal oxide
Showing 15 entries (1 t	o 15)			First Previous	1 Next Last		_	_
Term \$	Title	\$	Related to 🗘	Hit relevance	Find studies			First P
NPO 1694	PARTICLE SIZE.D90		PC GRANULOMETRY	65.103	by endpoint	\$	Related to ≎	Hit releva
<u>NPO 1694</u>	Core size		PC GRANULOMETRY	55.203	by endpoint		Related to	Therefore
<u>NPO 1617</u>							ENM 9000006	12.363
<u>NPO 1694</u>	PARTICLE SIZE		PC GRANULOMETRY	45.119	by endpoint			
<u>NPO 1915</u>	Intensity Mean Hydrodynamic Diameter		PC GRANULOMETRY	34.528	by endpoint		<u>NPO 1486</u>	12.363
NPO 1915	Number Mean Hydrodynamic Diameter		PC GRANULOMETRY	34.528	by endpoint		ND0 4540	40.060
<u>NPO 1800</u>							<u>NPO 1542</u>	12.363
<u>NPO 1915</u> NPO 1800	Volume Mean Hydrodynamic Diameter		PC GRANULOMETRY	34.528	by endpoint		NPO 1544	12.363
NPO 1915	Z-Average Hydrodynamic Diameter		PC GRANULOMETRY	23.676	by endpoint		<u>MO 1544</u>	12.505
NPO 1916			FC ORANOLOFILIRI	23.070	<u>by endpoint</u>		NPO 1548	12.043
	Average Length		PC GRANULOMETRY	10.044	by endpoint			
CHEMINF 000416	Density		PC GRANULOMETRY	10.044	by endpoint		NPO 1550	12.043
	Diameter		PC GRANULOMETRY	10.044	by endpoint			
NPO 1915	MASS MEDIAN AERODYNAMIC DIAMETER		PC GRANULOMETRY	10.044	by endpoint			
	MASS MEDIAN DIAMETER		PC GRANULOMETRY	10.044	by endpoint	5. A. A.		MIL
							1 (Mar 19)	A CANADA

### rch for "metal oxide"

Hit relevance

Search

Find studies

by substance type

by substance

by substance

by substance

by substance type

by substance type

type

type

type

M type metal oxide

Ontology supported free text search

UNDER DEVELOPMENT

Free text search 0 with query expansion 🖲 based on the eNanoMapper ontology and annotated database entries The results are sorted by hit relevance 0 The free text search finds ontology annotated database entries (e.g. protocols and endpoints in the second column).

The last column is a link leading to a list of studies

Search for "size"



2 February 2016

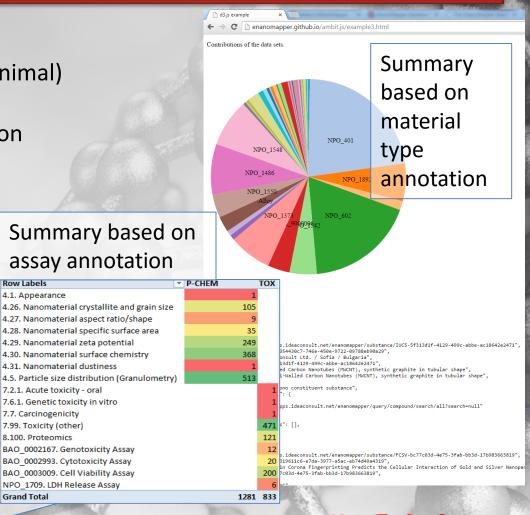
### The search quality depends on ontology annotation

### 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.



30



### Free text & faceted search (under development)

Nanomaterials
 OpenTox

#### ▲ > Search: >

#### Current Selection

#### remove all

(x) substanceType:NPO_1373 OR substanceType:NPO_1486 OR substanceType:NPO_1542

Search

(x) endpointcategory:BAO_0002167_SECTION_OR endpointcategory:BAO_0002993_SECTION_OR endpointcategory:PC_GRANULOMETRY_SECTION

#### Sources

#### NanoWiki P32 - KI

#### Nanomaterial type

Alloy Cattorhiarosanticle ENM 5000005 MatalOude NPO 1373 NPO 1486 NPO 1542 NPO 1544 NPO 1548 NPO 1550 NPO 1892 NPO 401 NPO 602 NPO 1403 NPO 5830 NPO 6330 NPO 660 NPO 6203 NPO 6403 NPO HO2 NPO 1403 NPO 5830 NPO 6400 NPO M030 NPO N203 NPO 1402 NPO 140 NPO 1402 NPO 140 NPO 140 NPO 1402 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NPO 1400 NPO 140 NP

#### Assay type

ASPECT RATIO SHAPE SECTION BAO 0002107 SECTION BAO 0002003 SECTION BAO 0003009 SECTION NPO 1709 SECTION PC GRANULOMETRY SECTION

#### SPECIFIC SURFACE AREA SECTION SURFACE CHEMISTRY SECTION

ZETA POTENTIAL SECTION

#### Protocols

DLS ELISA / TNF-α release in culture medium <u>Fpg-2Dmodified Comet Assay</u> <u>TEM</u>

#### Endpoint

DNA IN TBI MASS MEDIAN AERODYNAMIC DIAMETER PARTICLE SIZE

Results

#### References

http://dx.doi.org/10.1002/smll.201002366 http://dx.doi.org/10.1016/j.envpol.2008.11.004 http://dx.doi.org/10.1021/es9016975 http://dx.doi.org/10.1021/es9016975

http://dv.doi.org/10.1371/io

http://dx.doi.org/10.1021/nn2021056 http://dx.doi.org/10.1021/nn2021166 http://dx.doi.org/10.1021/nn3010087 http://dx.doi.org/10.1021/nn301622h http://dx.doi.org/10.1021/nn301622h http://dx.doi.org/10.1021/tx8002892 http://dx.doi.org/10.1021/tx8002892 Search

PARTICLE SIZE = 34 more

SPECIFIC SURFACE AREA = 250 more

[quest] Log out

< 1 2 3 4 5 > displaying 1 to 20 of 86

Help •

#### NM-111 [JRC2011 NM-111]

NPO 1542zinc oxide nanoparticle

- NPO_1709. LDH Release Assay (<u>\$)</u> [6] % cell viability = 102.057976 more
   BAO_0002993. Cytotoxicity Assay [20]
   TNF (ng/ml) = 13.771562 more
- 4.5. Particle size distribution (Granulometry) (S) [513]

#### 4.28. Nanomaterial specific surface area (<u>S</u>) [35]

#### SiO2 [Zhang2013 M17]

NPO 1373 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

PARTICLE SIZE = 13.5 more

#### SiO2-200 [Rancan2012 M3]

NPO 1373 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

PARTICLE SIZE = 190 more

#### ASP30F [Docter2014 M4]

NPO 1373 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

PARTICLE SIZE = 30.6 more

#### TiO2 [Shi2012 NM1]

NPO 1486 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

PARTICLE SIZE = 29.2 more

#### ZnO [Kim2012 NM2]

NPO 1542 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

PARTICLE SIZE = 50 more

#### NM-110 [JRC2011 NM-110]

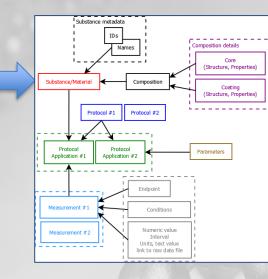
NPO 1542 nanomaterial PC GRANULOMETRY SECTION PARTICLE SIZE study

### Comments welcome at

https://github.com/enanomapper/data.enanomapper.net/issues/8

### **Semantic search**

Multiple Input formats



**Triple store** 

W3C 💗 SPARQL

# BioAssay Ontology

### Semantic formats RDF/XML N3 JSON LD



2/2/2016

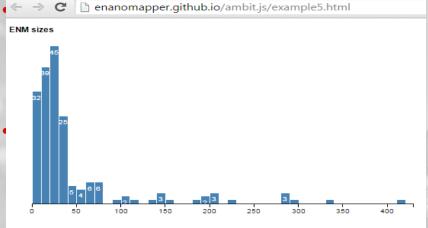
### **Tailored interfaces**

- 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



#### **J SON** reply

{ 
"study": [
{ "uuid": "NWKI-03949a0e-b3b6-46d1-8669-b5cdde29b34e", "owner": {
"substance": {
"uuid": "NWKI-db562080-5286-327a-b813-6775c437385e"
h
"company": {
"uuid": "NWKI-9f4e86d0-c85d-3e83-8249-a856659087da",
"name": "NanoWiki"
3
}•
"citation": {
"title": "http://dx.doi.org/10.1073/pnas.0802878105",
"year": "0",
"owner": "PNAS"
},
"protocol": {
"topcategory": "P-CHEM",
"category": {
"code": "PC GRANULOMETRY SECTION",
"title": "4.5 Particle size distribution (Granulometry)"
Ъ
"endpoint": "Primary Particle Size",

 – eNanoMapper database is based on the open source project
 2/2/2016 <u>http://ambit.sf.net</u>

ENM 🛞 🛞 🛞 https://apps.ideaconsult.net/enanomapper/api-docs	api_key	Explore								
eNanoMapper prototype database API AMBIT REST web services 2.7.2 [with enanomapper profile]. More at <u>https://apps.ideaconsult.net/enanomapper</u> Terms of service <u>Contact the developer</u> License										
algorithm : OpenTox Algorithms service	Show/Hide List Operations Expand	Operations Raw								
bundle : Datasets of substances	Show/Hide List Operations Expand	Operations Raw								
compound : OpenTox Chemical Compounds service	Show/Hide List Operations Expand	Operations Raw								
dataset : OpenTox Dataset service	Show/Hide List Operations Expand	Operations Raw								
feature : OpenTox Feature service	Show/Hide List Operations Expand	Operations Raw								
model : OpenTox Prediction Models service	Show/Hide List Operations Expand	Operations Raw								
property : Chemical substances Properties service	Show/Hide List Operations Expand	Operations Raw								
query : Queries	Show/Hide List Operations Expand	Operations Raw								
compound : Chemical structures search	Show/Hide List Operations Expand	Operations Raw								
substance : Substance search	Show/Hide List Operations Expand	Operations Raw								
substance : Chemical Substances service	Show/Hide List Operations Expand	Operations Raw								
GET /substance		List substances								
POST /Substance	Import substance(s) and studies									
GET /substance/{uuid}	Get a substance									
GET /substance/{uuid}/composition	Get substance composition									
GET /substance/{uuid}/structures	Get substance composition as a dataset									
GET /substance/{uuid}/study	2/{uuid}/study Get substance study									
Get         /substance/{uuid}/studysummary         Get study summary for the substance										
substanceowner : Substance owners	Show/Hide List Operations Expand	Operations Raw								
task : OpenTox Task service (asynchronous jobs)	Show/Hide List Operations Expand	Operations Raw								

[ BASE URL: https://apps.ideaconsult.net/enanomapper/api-docs , API VERSION: 2.7.2

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

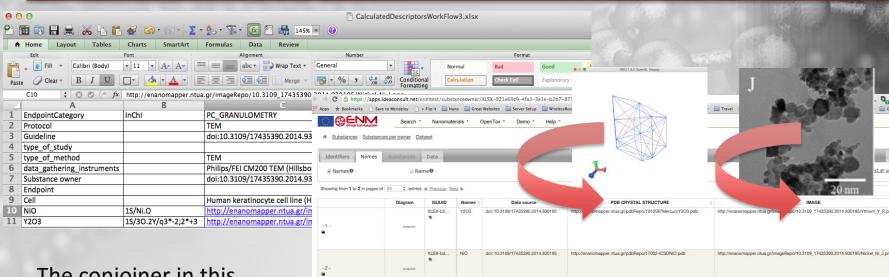


# WP4: Descriptor calculations (uses API to retrieve DB data and store back results)

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

00110000	···P									broki av	CONTRACT NAMES		
	-	- → C 🔒 http	://apps.ideacons	ult.net/enmtest/ui/_datas	et?dataset_uri=	https%3A%2F%2	Fapps.ideaconsult.n	et%2Fenmtest%2	Fdataset%2F399	63Ffeature_uris%5B%5D%3Dht	ttps%253A%252F%252Fa	p ☆ 📚 🗿 🔒	
	· .									f Facebook 🈏 Twitter 🚞 Ti			
	Apps				- oreat website	Jerrer Setup		bensor rep	pic elettud			metabolomics w miciped	
	Image			Search   Nanomate	rials 🔻 Op	oenTox 🔻 D	emo 🔹 Help 🔹						
		Home > Datase	s > Dataset brows	ser > https://apps.ideacor	nsult.net/enmte	st/dataset/39?fe	eature uris						
		Identifiers	Names Ca	alculated Other	_		_	_	_		_		
	1	🗹 NO. OF FI		🗹 TOTAL ENERG	Υ <b>θ</b>	🗹 FINAL	HEAT OF FORMAT		IZATION POTE	NTIAL [®] <b></b> ■ ELECTRO	NIC ENERGY		
		_		_				-		0			
Before filter applic	a	COKE-CO	RE REPULSION	MOLECULAR V	VEIGHT	S EHOMO		🗹 ELUI	400				
		Showing from 1 to	Showing from 1 to 1 in pages of 20 + entries + Previous Next Filter										
				Diagram	NO. OF FI	ILLED LEVELS $_{eig}$	TOTAL ENERGY	FINAL HEAT O	F FORMATION	IONIZATION POTENTIAL	ELECTRONIC ENERG	Y + CORE- CORE REPULSION	•
	Averag 1 2 3 4 5	- 1 - F			88		-7060.669	1672.661		3.898	-59148.238	52087.566	
	6												
	8												
	9		626	100	92	2.083	0.928		Minor				
	10		790	100		03.983	0.918		Major	2775			
	11		787	100		)5.397	0.890			ed Density			
	12		969	100	11	5.640	0.911		Vurtosis		NanoMa	PPP	

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



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.jagpot.org:8080/jagpot/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,





IMAGE

390 2014 930195/Vttrium Y P nn

# 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.





### Chemical /Toxicogenomics DB (no explicit NM support)

Pub hem







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

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

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

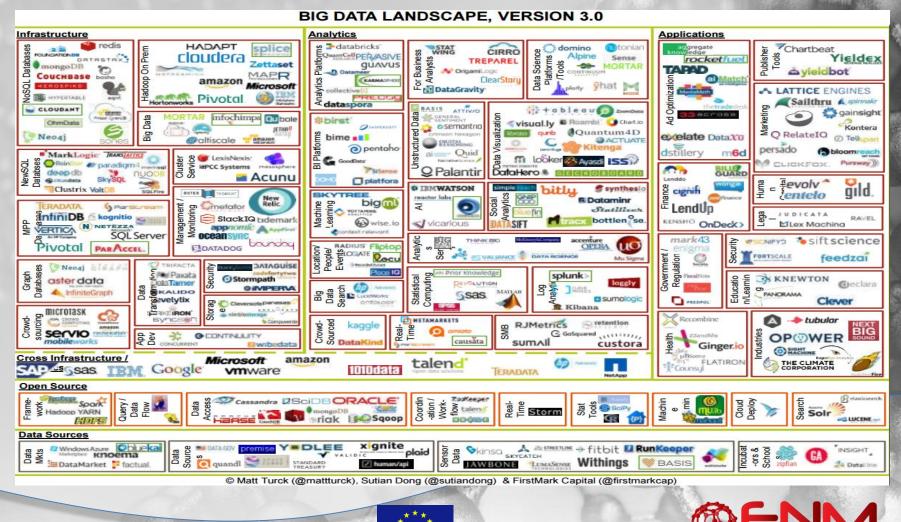
Comparative Toxicogenomics Database Includes nanomaterial related data. *HTML* 

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"



# 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





# **TOOLS AVAILABLE**





2 February 2016

# data.enanomapper.net

Nanomaterial bundles Dov	wnload	Help	_		This is a demonstration.	
Title 👻	Version \$	Code \$	Status \$		The data from the community	
Protein Corona Fingerprinting Predicts the Cellular Interaction of Gold and Silver Nanoparticles	1	10.1021/nn406018q	dente P	protein Co dataso	reconciliation of multiple terms	
OECD Harmonized Templates import test	1	Multi-Walled Carbon Nanotubes (MWCNT), synthetic graphite in tubular shape	br.	MWCNT	providing search and supporting data analysis	
NanoWiki	<u>1</u>	NanoWiki	pub	NANOWI		
In Vitro data from FP7 MARINA project (KI)	1	Comprehensive In Vitro Toxicity Testing of a Panel of Representative Oxide Nanomaterials: First Steps towards an Intelligent Testing Strategy	published	enanomap	<u>a8e4-20564e7eada7</u>	





### 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



### Format conversions, ISA schema

and manufacture discourse in the

[US] https://github.com/enanomapper/nmdataparser

The **nmdataparser** Java library is a configurable parser allowing to importing spreadsheet substance composition, characterisation and assay data into the eNanoMapper database, via [API]. The parser converts the spreadsheet into the internal AMBIT data model, using a JSON file for mapping the objects. The main class <code>GenericExcelParser</code> iterates over entries of *.xls and *.xlsx files returning a set of <code>SubstanceRecords</code> objects.

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. doi:10.3762/bjnano.6.165

Support for ISA-TAB / ISA-JSON pending

Stable release DOI 10.5281/zenodo.34065

<dependency>
 <groupId>net.enanomapper</groupId>
 <artifactId>nmparser</artifactId>
 <version>1.0.0</version>
</dependency>

Open source, available at GitHub





### Downloadable

#### 📄 eNanoMapper database | 🗙

→ C 🗋 ambit.sourceforge.net/enanomapper.html

### ambit

Home Downloads Documentation Publications Support

### The eNanoMapper prototype database

is part of the computational infrastructure for toxicological data management of engineered nanomaterials, developed within the EU FP7 eNanoMapper project.

- Provides support for upload, search and retrieval of nanomaterials and experimental data through a REST web services API and a web browser interface.
- Implemented by a customized version of AMBIT web services.

# Download and install

The eNanoMapper prototype database is an open source web application, which can be downloaded, installed and hosted by individual researchers or labs, and as such presents an open distributed platform for nanomaterials data management.

Download

Install guide

Feedback

### **Publications**

- N. Jeliazkova et al., The eNanoMapper database for nanomaterial safety information, Beilstein J. Nanotechnol., vol. 6, pp. 1609–1634, Jul. 2015.
   10.3762/bjnano.6.165
- 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. doi:10.1109/BIBM.2014.6999367
- N. Kochev, R. Grafstrom, N. Jeliazkova How to store nanomaterial safety data: meet eNanoMapper database, poster at SENN 2015

#### About



Nîna 🗖 📈 🗙

### 2 February 2016

**Quick Start** 

## 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.

ilstein-journals.org/bjnano/single/articleFullText.htm?publicId=2190-4286-6-165

About | My Journal | Support & Contact | Terms & Conditions



×

HOM

Then

All Vo Lates

Most Autho

GO 1

Volur No./F

EXC

Má

ZI

This prese contr

scien

### BEILSTEIN JOURNAL OF NANOTECHNOLOGY

E S	SUBMISSION	HELP				ADVANCED SEARCH			
NSE		•							
iatic Se Iumes	ries	-							
t Article Access			The eNanoMapper dat	tabase for nanomateri	al safety information				
ors			Barry Hardy ⁴ , Janna Ha	stings ⁵ , Markus Hegi ⁴ , V , Haralambos Sarimveis	Philip Doganis ² , Bengt Fade edrin Jeliazkov ¹ , Nikolay Koo , Bart Smeets ⁷ , Pantelis Sopa	hev ^{1,6} , Pekka Kohonen ³ ,			
	TICLE - select - ▼ - select - ▼	]	¹ Ideaconsult Ltd., Sofia, Bulgaria ² National Technical University of Athens, School of Chemical Engineering, Athens, Greece ³ Karolinska Institutet, Stockholm, Sweden ⁴ Douglas Connect GmbH, Zeiningen, Switzerland ⁵ European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom ⁶ Department of Analytical Chemistry and Computer Chemistry, University of Plovdiv, Plovdiv, Bulgaria ⁷ Department of Bioinformatics, NUTRIM, Maastricht University, Maastricht, The Netherlands ⁸ Computer Science Faculty, University of A Coruna, A Coruña, Spain ⁹ INT Institute for Advanced Studies Lucca, Lucca, Italy						
USIV	E	- 11		h (IST), Basel, Switzerlan					
iga					rmatics for environmental health	n and biomedicine".			
e	Beilstein Magazir	ne	Beilstein J. Nanotechnol.	2015, 6, 1609–1634.		doi:10.3762/bjnano.6.165			
nts ess	cess publication says and other from leading	-	Received 31 Mar 2015	Accepted 03 Jul 2015	Published 27 Jul 2015	Full Research Paper			
			Abstract						

Background: The NanoSafety Cluster, a cluster of projects funded by the European Commision, identified the need for a computational infrastructure for toxic ological data management of engineered nanomaterials (ENMs). Ontologies, open standards, and interoperable designs were envisioned to empower a harmonized approach to European research in nanotechnology. This setting provides a number of opportunities and challenges in the



# More details on eNanoMapper?

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



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

2 February 2016

Questions?

# **THANK YOU!**





2 February 2016