How to store nanomaterial safety data : meet eNanoMapper database



a) University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry, 24 Tsar Assen Str., 4000 Plovdiv, Bulgaria b) Institute of Environmental Medicine, Karolinska Institutet, P.O. Box 210, SE-171 77 Stockholm, Sweden c) Ideaconsult Ltd, 4 Angel Kanchev Str., 1000 Sofia, Bulgaria

* To whom correspondence should be addressed. e-mail: jeliazkova.nina@gmail.com, twitter: @10705013



http://enanomapper.net

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

eNanoMapper FP7 project

Objectives:

IQ_FA

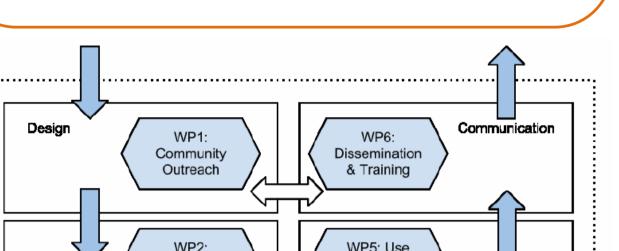
• Develop an ontology and database unifying information about nanomaterial safety

Karolinska

- Cover the full lifecycle from manufacturing • to environmental decay or accumulation
- Pan-European project, 8 partners \bullet
- Ontology growth through community and re-use

Vision:

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

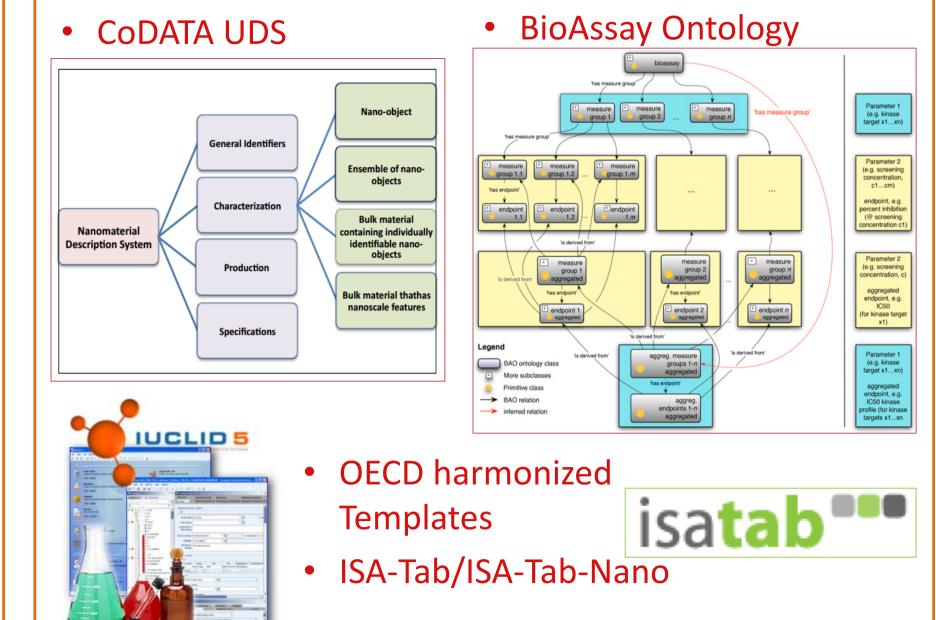


Chemical structures, substances, nanomaterials

Substance definition:

- NanoParticle Ontology (NPO): a Nanomaterial (NPO_199) is an equivalent class to chemical substance (NPO_1973) one of (nanoobject, nanoparticle, engineered nanomaterial, nanostructured material, nanoparticle formulation). The chemical substance itself is a subclass of a chemical entity (NPO_1972).
- DOI: 10.1007/s11051-013-1455-2 , . J. Nanoparticle Res. 2013, 15 : Compares the definition of the terms "substance" and "material" are in ISO, REACH and general science definitions of the terms. The paper notes the OECD HT definition of "reference substances" is very similar to the definition of the term "reference material".
- **REACH Guidance:** "Chemical substance, a material with a definite

Existing data models



Challenges

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

• Physicochemical identity

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

• Biological identity

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

• Data formats, Provenance, Visualisation

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

• Support for data analysis

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

WP2: Ontology WP5: Use Case Development Development WP3. Database WP4: Analysis development and & Modeling plementatio Application

WP7: Management

chemical composition". The definition of a substance encompasses all forms of substances and materials on the market, including nanomaterials; and may have complex composition.

Measurements

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

Materials sample

– Protocols, protocol

Experimental conditions

• Measurement groups,

• Raw data, derived data

ch substances by study protocol paramete

Search endpoint summa

Classification = Anionic

Classification = Cationi

Also contained

Also contained <u>in...</u>

Also contained

Structure

• Measurements,

parameters

Readouts

• Commonalities:

- Core – Coating(s)
- Linkage

Nanomaterials

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



Supported import formats

	nonized Templates,
IUCLID5	🚊 🖓 🌜 4.24 Agglomeration/aggregation
	Agglomeration/aggregation.001
	🗄 🖷 🐑 4.25 Crystalline phase
Navigation Codpoint study record. Applomeration/apprepation.001 Overv results Felders Section tree Detail level Administrative Data Data Source Materials and methods	
Complete 🔍 🚽 all fields 💗 Results and discussions Overall remarks, attachments Applicant's summary and conclu	🗄 💮 🛸 4.26 Crystallite and grain size
Appionerste/Appresste diameter	
	🗄 🐨 🛸 4.27 Aspect ratio/shape
🖗 👒 2 Classification & Labelling and PBT assessment	
	🗄 🗣 🎭 4.28 Specific surface area
41 Appearance/physical state/colour	🚛 👘 🚛 4.20 specific surface area
A 2 Metting point/freezing point St. Deviation 5 nm	
A 3 Boling point pH > ¥ 3	표 😪 4.29 Zeta potential
- 🙀 4.5 Particle size distribution (Granulom etry) Medium xxx %	
4.6 Vapour pressure Pemarks G	🕀 🐑 🐑 4.30 Surface chemistry
A Billion Solution	👜 🥐 4.50 Surface circuitstry
4.9 Solubility in organic solvents / fat solubil Apglomerate/Aggregate size	📩 📫 A 21 Duetingen
4.10 Surface tension 4.13 Flash point	🖽 🖓 🗭 4.31 Dustiness
- 🖗 4.12 Auto filammability	
A 13 Fianmability Percentile 050 G	🗄 🖷 🥯 4.32 Porosity
◆ 4.14 Depresenties Mean > ♥ 33 < ♥ 55 nm %	· · · · · · · · · · · · · · · · · ·
116 Oxidation reduction potential	🗄 🖷 📚 4.33 Pour density
417 Stability in organic solvents and identity St. Deviation 5 St. Deviation 5 St. Deviation 5	🔤 🖙 🖝 4.55 Four density

C 🔒 http	ps://apps.idea	consult.net/enanom	apper/ui			ୟ 🏠 P	• =
<u>=NM</u>	Search •	Nanomaterials 🔻	OpenTox 🔻	Demo 🔻	Help 🔻		<u>Log in</u>
Wel	come t	o eNanol	lanner	proto	type	databa	926
		O eNano e-by-design engineered nano mate		proto	type	databa	ise
				proto	type	databa	ise
A substance d				proto	type (databa	ise
A substance d	latabase to support safe	e-by-design engineered nano mate		proto	type (databa	ise

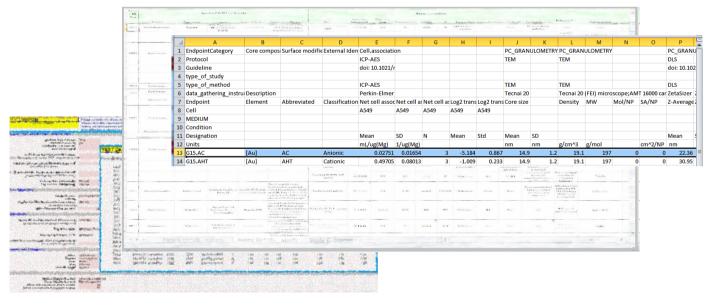
Advanced: Structure search | Search nanomaterials by identifiers | Search nanomaterials by endpoint data

eNanoMapper prototype database <u>http://data.enanomapper.net</u>

Physicochemical and toxicity data	Search API				
Search Vanomaterials OpenTox Demo Help V	compound : Chemical structures search Show/Hide List Operations Raw GET /query/compound/{term}/{representation} Exact compound search				
<u>Substance</u> NWKI-9f37da26-8619-3eb1-9c29-e5f9ea09de54 Study	GET /query/similarity Similarity search GET /query/smarts Substructure search				
IUC Substance Composition P-Chem (4) Tox (5) Filter Expand all Collapse all	substance : Substance search Show/Hide List Operations Expand Operations Raw				
Expand all Collapse all	GET /query/substance/facet Search substances by study owner GET /query/substance/reference Search substances by reference structures				
I.5 Particle size distribution (Granulometry) (2)	GET /query/substance/related Search substances by related structures GET /query/substance/study/experiment/{term} Search substances by protocol application parameters				
Distrib Distrib <thdistrib< th=""> <thdistrib< th=""> <thdistrib< th=""></thdistrib<></thdistrib<></thdistrib<>	GET /query/substance/study/owner/{term} Search substances by study owner				

A 12 Storing in station, your reaction in sources A 12 Storing in station, your reaction in sources A 12 Storing in station in sources in a station in the story of the story	Minimum Image: Contract of the image: Contrac
⊕ 4.25 Sorfac Censistry ₩ 9.43 Sorfac Censistry ₩ 9.43 Devisiting ₩ 9.43 Provisity ₩ 9.43 Processing to the tensistry ₩ 9.43 Processing form store potential ₩ 9.43 Censistry	Distribution
11 Guardon on saf viet 12 Guardon sarch 13 Assessment reports 14 Information requirements	Mature Service Service Service

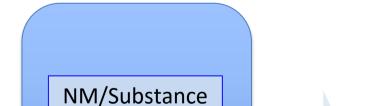
- NanoWiki RDF (KI, MU)
- Custom spreadsheets (NanoSafety cluster)

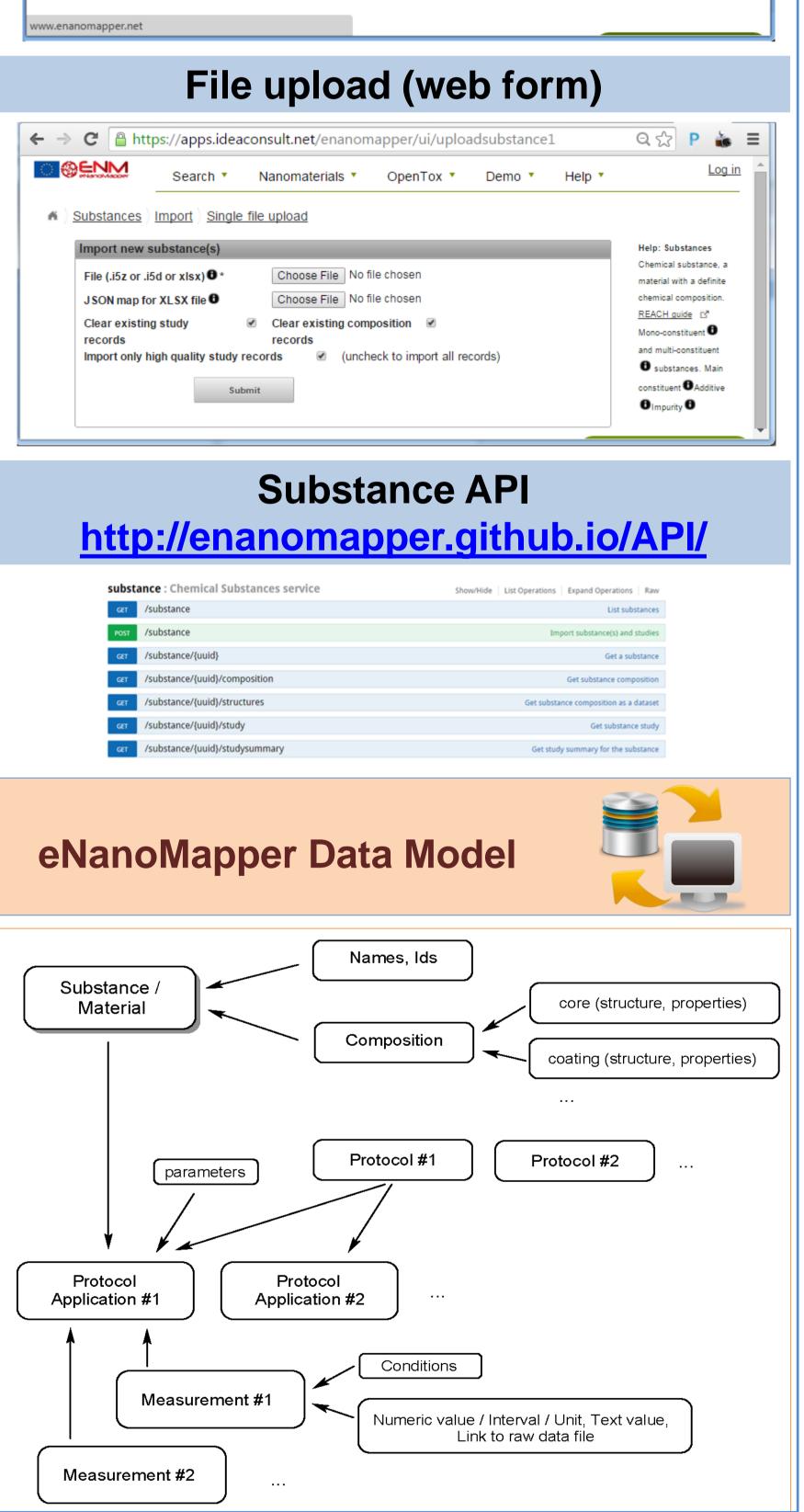




(under development)

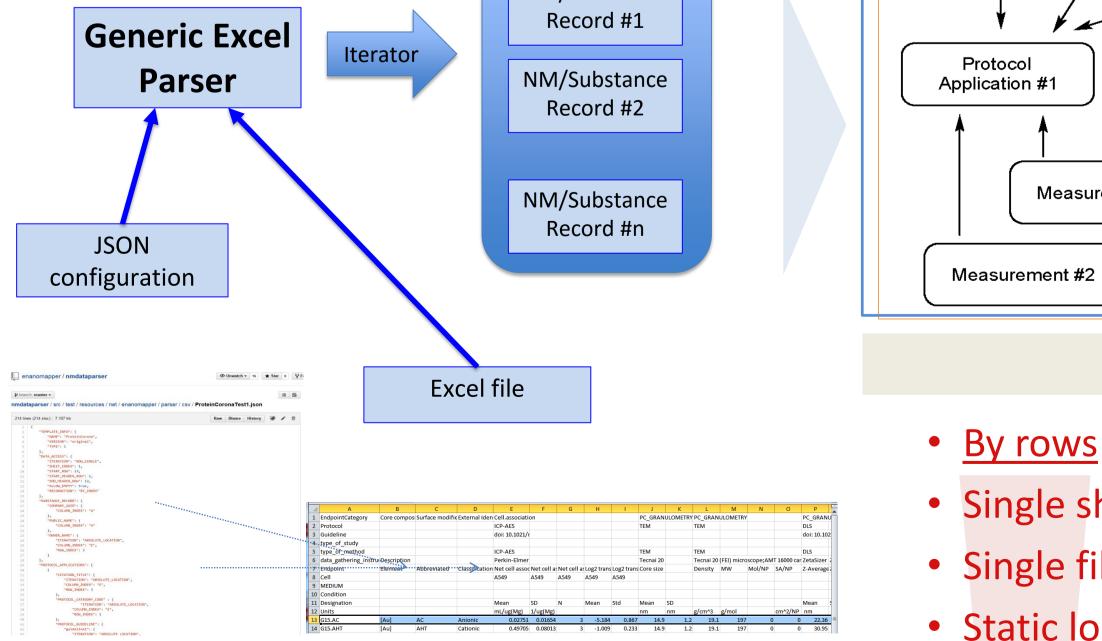
Configurable Excel Parser for custom spreadsheets





Micron												uery/substance/refe				s
4.5 Particle size distributi	on (Granulome	try) (2)										uery/substance/refe				
Tort													y/experiment/{term}		Search	h sub:
Test DistribPassag Materia _ type num.	Endpoint	♦ Value	Reference	Gu	ideline 🕴	Method type	Owner	\$	U	UID 🔶	GET /q	uery/substance/stud	y/owner/{term}			
Form						-//			WKT-24220	59f-8955-48b0	GET /q	uery/substance/stud	y/protocol/{term}		2	Searc
	PARTICLE SIZ	E = 221	DOI		DLS	DLS	-		ð	591-6955-4600	GET /Q	uery/study				
	PARTICLE SIZ	E = 221	DOI			-	-	1	WKI-bc3b5b	948-5780-401e						
Showing 2 study(s) (1 to 2)										EN	Л Со	mpo	nent	S	
IUC Substance Cor	nposition	P-Chem (4) Tox (5)		_		-			_			-			
licron								<u> </u>		Search •	Nanomaterials <	OpenTox •	Demo 🔻 Help	•		
3.100 Cell Viability Assay (5)							▲	ch substances	<u>by identifiers</u>						
Reference	Cell line	Doses/concentrations	Endpoint 🗍	Result	Result (text)	Cuidaliu	Owner	Showing from	n 1 to 5 in pages	of 10 ▼ substan	ces 📹 <u>Previous</u> <u>Ne</u>				Filter	
Reference		Dosesy concentrations	Endpoint	Result	Result (text	Jourdein	Owner		Substance	Substance UUID	Substance Ty	pe	Reference substance UUID		Owner	
<u>2011</u>	HaCaT	= 100 mg/L	Percentage Viable Cells	= 95	-		Chemosph	□ -1- ~	G15.AC	FCSV-bc77c03d-4	nanoparticle	G15.AC	FCSV-50cca421-d	Protein Corona Finger		
http://dx.doi.org 2011	J/10.1016/j.chemosp HaCaT	ohere.2011.04.067 = 500 mg/L	Percentage Viable Cells	= 98	-		Chemosph	-		'⊐ Composition name: Composition UUID:	FCSV-bc77c03d-4e	75-3fab-bb3d-17b98366	3819	Interaction of Gold and	D Silver Nanoparticles	<u>.csv</u>
2011	HaCaT	= 1000 mg/L	Percentage Viable Cells	= 92	_		Chemosph		Puri _{Type}	ty of IUC Substance:	T EC No		Typical concentra	tion 🔺 Conce	ntration ranges	\$
	Hucur	_ 1000 mg/L	l'ercentage viable cens	- 52			chemosph		Core	₽ [Au]			0 % (w/w)	0 % (w/w)	0 % (w/w)	
<u>2011</u>	HaCaT	= 10 mg/L	Percentage Viable Cells	= 101	-		Chemosph									
2011	HaCaT	= 7000 mg/L	Percentage Viable Cells	= 74.3	-		Chemosph									
owing 5 study(s) (1 to 5)								Coating	e (2r)-2-Acetamido-	3-Sulfanyl-		0 % (w/w)	0 % (w/w)	0 % (w/w)	
										Acid, Pwkskimoespyia Bypyzuonsa-						
		Doto of	allaction							N,Inchi=1s/C5h9no3s 4(2-10)5(8)9/H4,10h,2						
		Data co	ollection	15						(H.6,7)(H.8,9)/T4-/M0 Acetamido-3-Sulfanyl						
		ъ.	х							Acid,(2r)-2-Acetamido Mercapto-Propionic A						
$- \rightarrow \mathbf{C}$ \triangleq https:/	/apps.ideac	onsult.net/enanoma	pper/bundle?page=0&p	bagesize=	10	Q ☆	🍓 ≡			Acetamido-3- Mercaptopropanoic A	cid.N-					
<mark>⊚ENM</mark>	Search 🔹	Nanomaterials <	OpenTox • Demo •	Help 🔻	,		Log in			Acetyl-L-Cysteine					Search	h:
			-					- 2 -	G15.AHT	FCSV-8f5cd32a-3	nanoparticle	G15.AHT	FCSV-50cca421-d	Protein Corona Finger	printing Predicts the C	Cellul
 <u>Dataset of substa</u> 	inces and stu	dies								Composition name:			6	Interaction of Gold and	d Silver Nanoparticles	<u>.csv</u>
									Puri	Composition UUID: ty of IUC Substance:	FCSV-8f5cd32a-33	50-300b-91d0-87000ee5	d7ee			
Showing 3 substance da	tasets (1 to 3)			First P	revious 1 N	lext La	st		Туре	▼ Name	EC No		Typical concentration	· · · · · · · · · · · · · · · · · · ·	ntration ranges	•
*		Title			-	Downlo	ad		Core	6 [Au]			0 % (w/w)	0 % (w/w)	0 % (w/w)	
★5	lataset#3	1100				-										
Protein Corona	Fingerprinting		eraction of Gold and Silver Na		Protein Co datase				Coating	of 6-Amino-1-Hexan	ethiol		0 % (w/w)	0 % (w/w)	0 % (w/w)	
Source rights	aures Substar	ices oubstances and sti	<u>udies Properties list Summa</u>	iy <u>wennig</u>					ovally	•			0 % (WW)	0 % (WW)	0 % (WW)	
★5 <i>e</i> <u>Substance</u>						🖻 🚳										
OECD Harmon Metadata Strue			<u>udies Properties list Summa</u>	ry Workina	mating MW	CNT										
Source rights												1	1	1	Searc	
★5	lataset#1					© 7		🗆 - 3 -	G15.Ala-SH	<u>FCSV-9cc7dd73-d</u> ⋐	nanoparticle	G15.Ala- SH	<u>FCSV-50cca421-d</u> ®	Protein Corona Finger	-	
	tures <u>Substar</u>	nces Substances and stu	udies Properties list Summa	ry Working	NANO	WIKI		🗆 - 4 -	G15.Asn-SH	FCSV-5b29295c-f	nanoparticle	G15.Asn- SH	FCSV-50cca421-d	Protein Corona Finger		
rights							ļ	□ - 5 - ■	G15.AUT	FCSV-f3d54c07-7	nanoparticle	G15.AUT	FCSV-50cca421-d	Protein Corona Finger	printing Predicts the C	Cellul
										5			<u>ت</u>			
naro th		INA cha	racterisa	tion		1						viatr	ix vi	ew		
pare tr			i acterizo		Ian	L		Identifiers								

🔲 15UUID 🕒 🗹 Contained As 🕄



https://github.com/enanomapper/nmdataparser

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

Data access scenarios

- <u>By rows</u> / by columns
- Single sheet / multi sheet
- Single file / multi file
- Static location definition / Dynamic location definition
 - ROW_SINGLE

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

• ROW MULTI DYNAMIC

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

ABSOLUTE LOCATION

Data is always taken from a particular cell

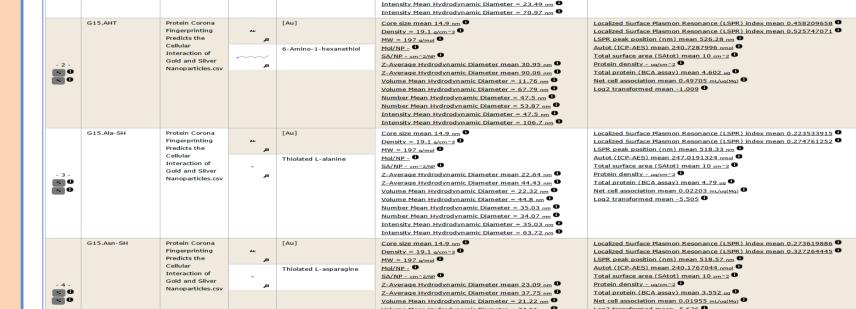
data model and upload into the DB;

input templates into the eNanoMapper

assay data using your own templates;

• The Excel parser enables converting the

- Search and explore the eNanoMapper DB;
- Format convertors by exporting the data model into different formats (under dev) **Ontology annotation (under development)**



Mapping the Excel layout into the data model by JSON configuration

