

5 – 9 June

🚲 Karolinska

# Automatic workflow for HTS data FAIRification, preprocessing and Tox5 in-vitro toxicity scoring



G. Tancheva<sup>a,b</sup>, P. Nymark<sup>c</sup>, V. Hongisto<sup>d</sup>, N. Kochev<sup>a,b</sup>, N. Jeliazkova<sup>a</sup>

<sup>a</sup> Ideaconsult, 4 Angel Kanchev St, 1000 Sofia, Bulgaria

<sup>b</sup> University of Plovdiv, Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, 24 Tsar Assen St, 4000 Plovdiv, Bulgaria

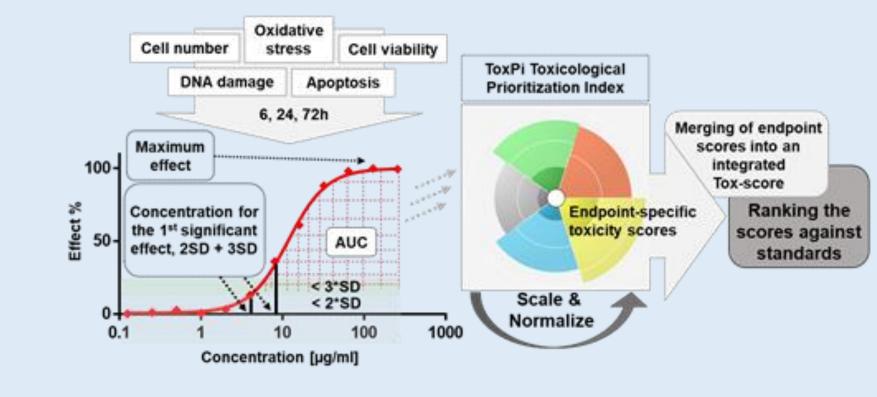
<sup>c</sup>Karolinska Insitutet, Institute of Environmental Medicine, 171 77, Stockholm, Sweden

<sup>d</sup>Misvik biology, Division of Toxicology, Karjakatu 35 B, 20520, Turku, Finland



## **BACKGROUND MOTIVATION AND OBJECTIVE**

- New chemical substances and materials, including nanomaterials, pose safety challenges.
- Regulatory agencies are interested in using safety data generated through New Approach Methodologies (NAM).
- Data management based on FAIR guiding principles helps with consistent curation and reusing of accumulated data.
- Nanosafety, cheminformatics, and bioinformatics communities can benefit from this data management approach. • The generated high-throughput screening (HTS) biological data is used for efficient clustering, ranking, prioritization of NMs and read across.



Tox5 [1] in vitro toxicity scoring and ranking concept:

	A	1	в	C	D	E	F.		G	н	1	1.	J	к		ъ.					
1	PLATE R														_						
2						e															
3						Jun 06 202		8 GN	T+0300	Suome	norm	aaliaka	à								
4	User inter	face	program	n version	2.9	2.3								-							
5						02:			-			_									
6									-		1										
7											-										
8								-			-	_									
9	Temperat										1										
10	and the second second second					16:17:18:19:2	0-21-22-2	3.24				_			-						
11						83.253 188.2				525-261	290.20	1243-2	63.25	0.263	188						
12						82 260 281									5 C C C						
10	C:188.21	1.6.0	A	E			D	E	F		3	H			J	10.0	6	1			
14	D.25127	1	and the second				-	E.	-				-		4		÷.		-		
15	E:298.23	2				R-filter sh								-							
16	F:234:25	-				ART DATE: M				THE OWNER	00000										
	G:256:18					am version: 2				33 GPIT	03001:	suomer	nom	aasas	45						
17	A Date of the second													-			-				
18	H;274;25 1:306:216					vare version: version: 229				_	_										
19		6								-											
20		7				nber: 215000								-			-				
21		8				1EndPoint			-	-				-							
22	L:245:22					11, 12, 13, 14, 1		0.40.1	0.24.22	22.24											
23	M;325;21		1. C. C. C.		-	the second second second	and the second second	1000		and the second second		1 1 1 1 1 1	13.13		10.142	100					
24	N:290.22					08:176:105:1											-				
25	0,258,19		10-10-10-10-10-10-10-10-10-10-10-10-10-1		100, 10	01; 113; ION; IO	0,1/4,1/	0, 143	( 120; 10 (	100,103	100,114	, ION, IO	0, 10 /	1114(10	LUCK.	00					
26	P;283;24			159,171	1.4	A	в	С	0	1	E	F	1 3	G	н		1.	J		ĸ	L
27	PARAME			70,180;	1	PLATE RES	ULTS						1			-					
28	PROTOC		and the second s	178,168,	2	PROTOC SI	00ms #	R-filte	H shak	e:											
29	SAVE TI			155,153, 141,191,1	3	MEASUREM	ENT STA	ATO	ATE: Mo	n Jun Of	2022 1	0.18.03	GMT	+0300	(Suom	en nor	maalia	ika);;;			
30	PLATE: (		and the second se	458:26	4	<b>User interfac</b>	e progra	m ver	sion: 2.9	.2.3											
31	PLATE F			450,261	5	Communica	tion firms	are v	ersion: 2	.02;											
32	1. OPER/			55:169:	6	Instrument in															
33	Shake Ti			42,148	7	Instrument s	erial num	ber:2	2150006												
34	2. OPER			155,124;	8	2. OPERATI	ON: LUM	EndF	oint												
35	Z-FOCU			172:174	9	Temperature	e: 26.6;;;;														
36	No Mirror	23	100 March 100	156,153	10	;1,2;3;4;5;6;	7;8;9;10;	11, 12;	13, 14, 15,	16;17;18	19,20,2	1,22,23	3,24								
37	EMSFIL1			165, 160	11	A:109,153,15															
38	Meastim	25		157:171	12	B; 154; 171; 18	4,165,15	7;175	164:166	:174:178	143;125	5,161,16	5;183	158,17	4;194;	156, 18	7;174;	161,17	1,180		
39	PMT HV:	26		METERS	13	C; 120; 159; 17	71, 149, 12	0,176	144;206	152,164	133,13	5;135;1	39,13	9,231,2	86,318	236;	347;30	07;312	466,50	2	
40				DC 500	14	D:171:170;18	0:213:17	9,184	179,195	202.22	5,243,2	64:260	333	279,33	1,288,1	347.3	1069,1	971,168	64;1148	1048.54	8
		29	10 - C - C - C - C - C - C - C - C - C -	TIME: To	15	E; 160; 178; 16	8,177;	See.	A	в	C		D	E		F	0		н	1.1	
		30		C Flat	16	F;144;155;15	3, 136,	1 P	LATER						-			-			_
				FORM	17	G:140.141.15	1,140;		ROTOC						nanan						
		31		RATION	18	H;792;458;2	ALC: No. 1									18:03	GMT+	0300	Suome	noma	alaka
				Time: 0	19	153, 152, 17	5, 755, 1		lser inter												
		33	and the second second	RATION	20	J: 182; 155; 16	9,148,		ommuni								-				
				USImm	21	K;151;142;14	8,154,		atrumen								-				
					22	L:144:155:12	4.000		strumen								-				
		36		LTER 5	23	M; 148; 172; 1	74:154		OPERA												
		37		imelms:	24	N:145,156,15	53,166		emperat								_			-	-
			PMTH		25	0:153,165;1	60:162		12:3:4:5					7.18.15	20.21	22.2	3,24				
		-22	142011	v: 1000	26	P 156 157 17	11.156											272.1	77.994	12.172.1	72.540

HTS generates vast amounts of data, which are difficult to process and analyze. In addition, using the ToxPi software creates even greater difficulties:

requires time-consuming manual processing, b difficult to scale for larger NM

- normalization of the HTS metrics, separately in the range [0,1], for each time point and endpoint;
- combination of the normalized metric values to obtain final Tox5 endpoint scores.

	for the second second second	0.88.0	PR, MJC, DJC, DJC, DJC, MJC, HJC, NJC, DJC, DJC, DJC, PRJC, NJC, DJC, NJC, ZTJC, HTC, DRC, HJC, HJC, PRJC, PRJC, PJC, PDC, PJC, PJC, PJC, PJC, PJC, PJC, PJC, PJ	
27	PARAMETERS	12	B: 154: 171: 184: 165: 157: 175: 164: 166: 174: 178: 143: 125: 161: 165: 183: 156: 174: 194: 156: 187: 174: 161: 171: 180	
28	PROTOC 500ms	13	C 120, 153, 171, 149, 120, 176, 144, 206, 152, 164, 139, 195, 135, 139, 139, 231, 286, 318, 236, 347, 307, 312, 466, 502	
29	SAVE TIME: Tue A	34	D. 171 170 180 213 179 184 179 195 202 225 243 264 260 333 279 331 288 1347 3069 1971 1664 1148 1048 546	Ē
- 20	PLATE C Flat both	15	E 160, 178, 168, 177, 151, 185, 144, 217, 145, 139, 151, 181, 147, 165, 143, 172, 169, 208, 163, 192, 176, 142, 184, 166	
31	PLATE FORMAT: 1	16	F 144, 155, 153, 136, 190, 146, 173, 180, 166, 169, 158, 153, 177, 135, 146, 161, 166, 184, 172, 174, 166, 182, 132, 163	
32	1. OPERATION: OH	17	G. 140, 141, 191, 140, 165, 129, 169, 183, 139, 182, 172, 174, 138, 179, 171, 145, 172, 167, 166, 177, 185, 147, 172	
33	Shake Time: 0 hou	18	H 792:458:268:217.188:199.197.178:142.165.166.142.147.172.174.183.174.176.142.164.151.174.153.138	
34	2. OPERATION: LL	19	1 153, 152, 175, 155, 174, 147, 186, 154, 176, 145, 158, 160, 161, 152, 185, 160, 175, 178, 163, 173, 159, 169, 170, 168	
35	Z-FOCUS/mm: 8.5	20	1, 182, 155, 169, 148, 177, 148, 166, 180, 164, 141, 169, 159, 150, 175, 172, 174, 146, 159, 162, 170, 177, 178, 146, 184	
36	No Mirror,	21	K, 161, 142, 148, 154, 157, 146, 165, 145, 186, 180, 165, 168, 171, 168, 172, 136, 181, 151, 179, 184, 140, 172, 179, 151	
37	EMSFILTER: IRBk	22	L, 144, 155, 124, 158, 142, 148, 150, 139, 157, 148, 176, 154, 157, 152, 178, 172, 149, 183, 180, 173, 156, 159, 159, 166	
38	Meas time/ms: 500	23	M: 148, 172, 174, 154, 157, 143, 178, 177, 163, 151, 161, 160, 147, 203, 152, 183, 163, 166, 159, 160, 181, 158, 154, 156	
39	PMT HV: 1000 V Di	24	N: 145: 156: 153: 166: 156: 153: 143: 172: 170: 155: 156: 151: 160: 158: 130: 152: 148: 162: 161: 150: 147: 164: 166: 165	
40		25	0.153,165,160,162,166,173,150,160,150,158,132,154,153,164,152,132,158,161,136,155,176,167,172,172	
		26	P; 156, 157, 171, 156, 177, 147, 148, 137, 152, 152, 173, 166, 165, 145, 148, 148, 153, 161, 144, 178, 147, 143, 164, 183	
		27	PARAMETERS	
		28	PROTOC 500ms IR-filter shake:	
		29	SAVE TIME: Tue Apr 14 2020 13:41 20 GMT+0300 (Suomen normaaliaka)	
		30	PLATE: C Flat bottom:	
		31	PLATE FORMAT: 384 wells (16X24)	
		32	1. OPERATION: Others Shake	
		33	Shake Time: 0 hours 0 minutes 10 seconds Shaking in normal mode Shaking in circular form	
		34	2. OPERATION LUMEndPoint Top measurement	
		35	2-FOCUS/mm: 8.5:	
		36	No Minor	
		37	EMS FILTER: IR Blocker (400-650) EMS SPOT SIZE: 2 (Default)	
		-50	Maas timeling 500	

39 PMTHV: 1000 V Disor: 47;;;;;;

- datasets,
- $\blacktriangleright$  prone to errors
- challenging to implement FAIR principles

## METHODS

We have developed a new Python module for collection and annotation of raw data, consequent normalization and calculation of dose-response metrics. The module invokes ToxPi-R library and strictly follows the original Tox5 approach. The module can be used independently or as a part of developed by us Orange [2] workflow with custom widgets for fine tuning of the data processing.

Tox5

ð

lable Value

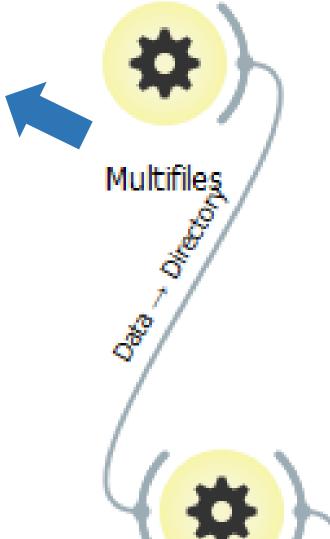
Tox5 view



### **Read HTS data**

- ✓ Read HTS meta data from annotation file
- ✓ Read HTS data from csv or excel formats

	🏶 Multifiles - Orange 🛛 🗆 🗙
	Multifiles
	Select files
	<ul> <li>Select directories</li> </ul>
	select directories
Tox5 calculation model	?   ⊡ 2 .::
✓ Cell selection	
✓ Automatic grouping by endpoint	or by time-endpoint



### **Toxpi preprocess**

- ✓ Custom data normalization
- ✓ Calculating dose response metrics

Endpoints CTG CASP DAPI H2AX 80HG	Normalization <ul> <li>Remove outliers by quantiles</li> <li>% of median control</li> <li>Subtract median blank samples</li> <li>Correct data by DAPI results</li> </ul>	$\geq$	Bit Mark & Status // St	NAME         NAME         NAME           NAME         NAME         NAME					
	Dose - response		2 4900 71.140 173.85 2 MCM5-05 21.807 71.752 4 Refuel_ 4 U.028 5 MCM5-19 4 4	material	A549-6H_AUC_DAPI	A549-24H_AUC_DAPI	A549-72H_AUC_DAPI	BEAS-2B-6H_AUC_DAPI	\S-2B-24H_
			6 MCM-06 81223 716.04 7 MCM-01 34.968 43.988 8 5-Fuerdiaci 4 57873	1 Mitomycin C	-0	122.905	308.365	-0	84.5813
	🗌 🔽 Calavilata daga 🛛 nagmanga								
	Calculate dose – response		9 000 -0 -0 10 RLOD-, -0 -0	2 4NQO	71.1447	173.685	262.075	158.043	276.574
	<ul> <li>Calculate dose – response</li> <li>parameters</li> </ul>		B         D(2)         -0         -0           10         PL-CD         -0         -0           11         Pum_Fel/OLAM.         -0         12.0865           12         NECME-21         -0         -0           12         Petes/01         4.1176         6.8668	2 4NQO 3 NRCWE-055	71.1447 31.887	173.685 78.7262	262.075 87.4904	158.043 18.2438	276.574 26.3497
			B         BQU         d         d           10         PLOD         4         4           11         Pum/(HOL)M         4         11.005           12         MEXH6-21         4         4						
-D - D -			9         0Q12         -0         -0           10         Prov. (+002, M.         -0         -0           11         Prov. (+002, M.         -0         10           12         Prov. (+002, M.         -0         -0           12         Prov. (+002, M.         -0         -0           15         Prov. (+002, M.         -0         -0           15         Prov. (+002, M.         -0         -0           15         Prov. (+002, M.         -0         -0           16         Prov. (+002, M.         -0         -0           17         Prov. (+002, M.         -0         -0           16         Prov. (+002, M.         -0         -0           17         Prov. (+002, M.         -0         -0           16         Prov. (+002, M.         -0         -0           17         Prov. (+002, M.         -0         -0         -0           16         Prov. (+002, M.         -0         -0         -0         -0           17         Prov. (+002, M.         -0         -0         -0         -0         -0         -0         -0         -0         -0         -0         -0         -0         -0	3 NRCWE-055	31.887	78.7262	87.4904	18.2438	26.3497

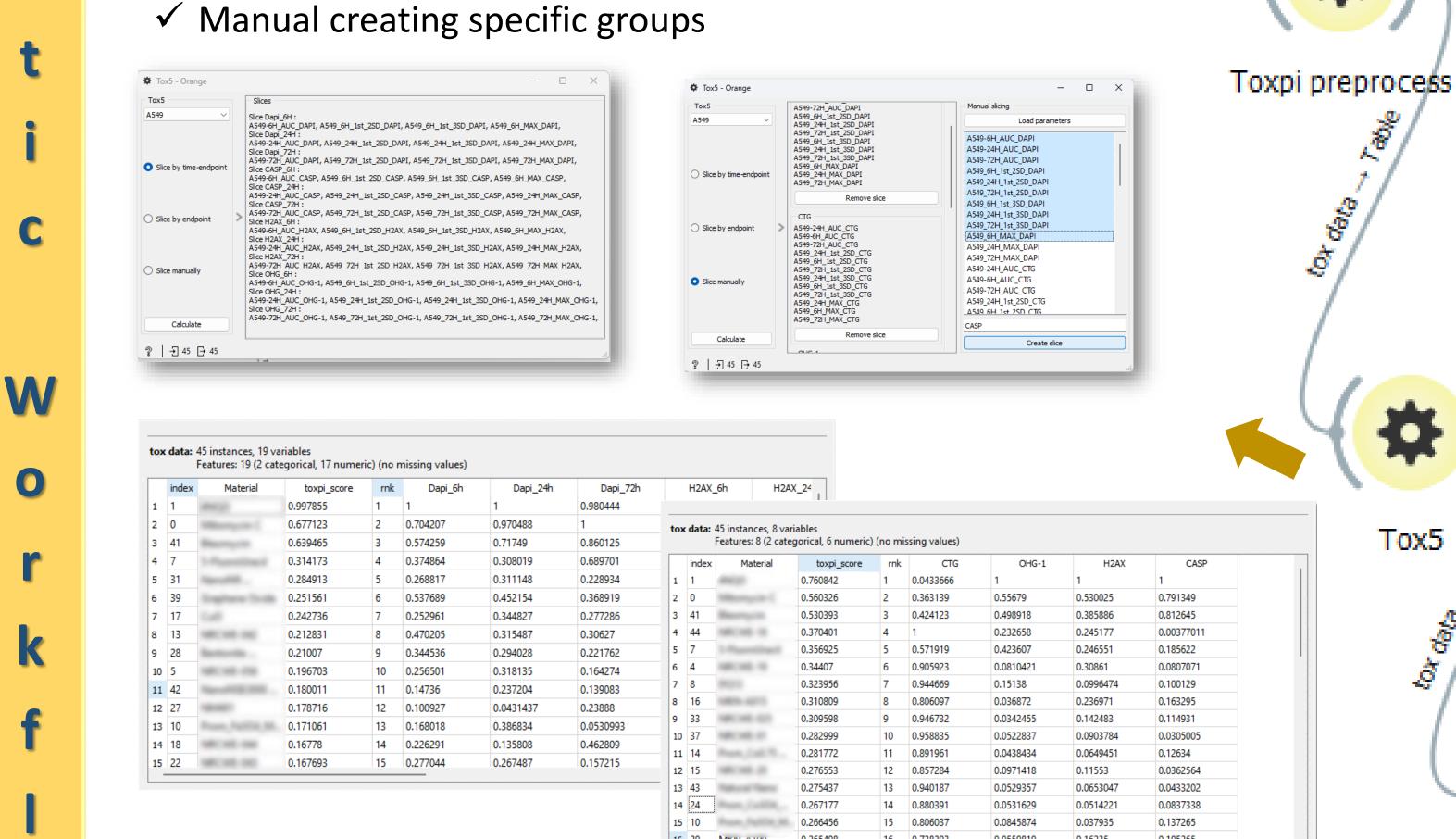


Table output with calculated TOX5 scores, ranks for different grouping

## FAIR data in eNanoMapper

eNanoMapper data model is flexible and

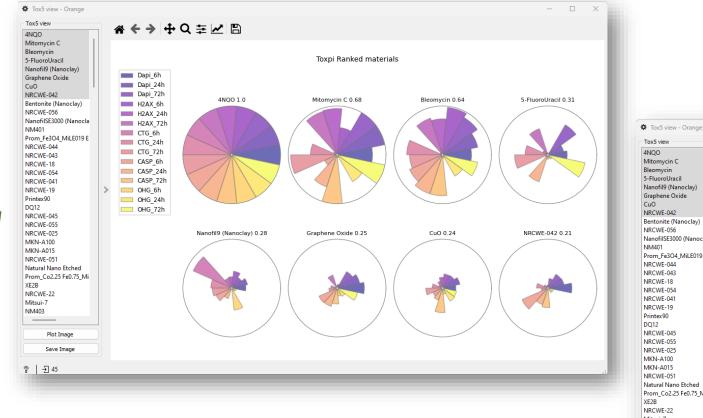
7	NRCWE-051	24.8663	40.3638	62.3096	29.6604	31.9798
8	5-FluoroUracil	-0	9.76613	154.067	-0	-0
9	DQ12	-0	-0	-0	-0	9.44054
10	PL-QD	-0	-0	-0	-0	-0
11	Prom_Fe3O4_M	-0	10.4965	-0	6.80965	-0
12	NRCWE-21	-0	-0	4.65025	12.3489	27.6729
13	Printex90	4.33706	6.84988	21.1908	40.7275	41.5641
14	NRCWE-042	135.436	117.605	112.014	35.5388	26.3168
15	Prom Co0.75	-0	-0	4.16853	8.22265	10.2659

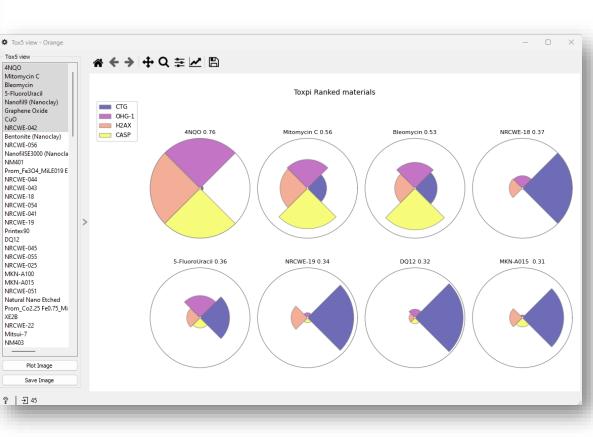
Table output with calculated dose-response metrics

#### **Tox5 pie view**

Y Pie view of ranks, and combined toxicity scores for each material

- $\checkmark$  User selectable set of materials for view
- ✓ Save image in convenient formats (e.g. png, jpeg)

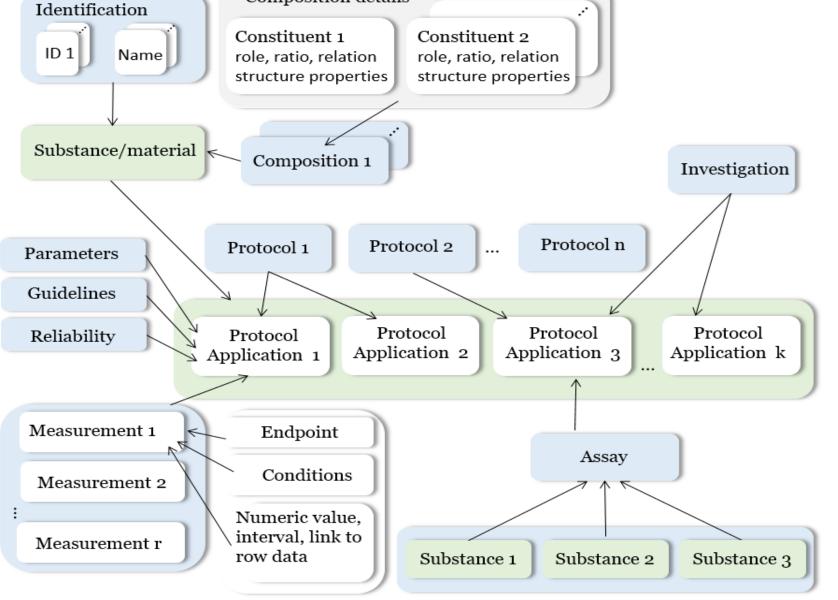






0

W



Composition details

The eNanoMapper database is an open source chemical substance data management solution, which currently possesses the largest searchable compilation of nanoEHS data in Europe [4].

dynamic and allows efficient FAIR support





Online user interface enabling user friendly access to the aggregated search index of (sub)set of eNanoMapper database instances [4].

- ✓ The new Python module enables faster data preprocessing calculations and eliminates the possibility of errors.
- ✓ Python module and Orange workflow are extending the eNanoMapper FAIRification workflow [3] by facilitating FAIRification of HTS data.
- ✓ The resulting FAIR data includes both raw and interpreted data (scores) in machine readable format.
- ✓ FAIR HTS data can be distributed as data archive and/or be integrated into eNanoMapper database [4].

1. Nymark, P; Hongisto, V et al. Toxicology Letters, 314, 2019, https://doi.org/10.1016/j.toxlet.2019.09.002 2. Demsar, J et al, Journal of Machine Learning Research, 2013, 2349–2353. 3. Kochev, N et al. *Nanomaterials*, *10*, **2020**, <u>https://doi.org/10.3390/nano10101908</u> 4. Jeliazkova, N et al. Nat. Nanotechnol. 16, 2021, 644–654 <u>https://doi.org/10.1038/s41565-021-00911-6</u>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 953183 HARMLESS.