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

G. Tancheva[a,b], P. Nymark[c], V. Hongisto[d], N. Kochev[a,b], N. Jeliazkova[a]

[a] *Ideaconsult, 4 Angel Kanchev St, 1000 Sofia, Bulgaria*

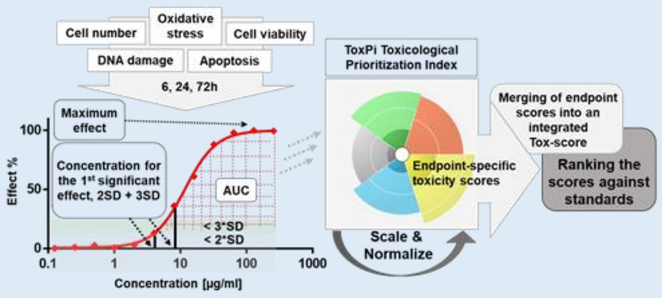
[b] *University of Plovdiv, Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, 24 Tsar Assen St, 4000 Plovdiv, Bulgaria*

[c] *Karolinska Insitutet, Institute of Environmental Medicine, 171 77, Stockholm, Sweden*

[d] Misvik biology, Division of Toxicology , Karjakatu 35 B, 20520, Turku, Finland

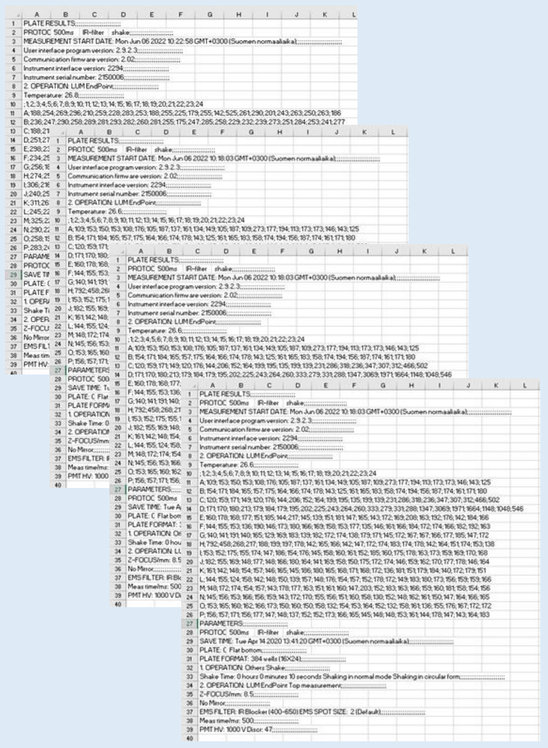
BACKGROUND MOTIVATION AND OBJECTIVE

- 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:

- 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.



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,
- ➢ difficult to scale for larger NM datasets,
- ➢ prone to errors
- ➢ challenging to implement FAIR principles

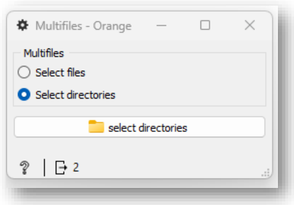
METHODS

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.

Automatic Workflow

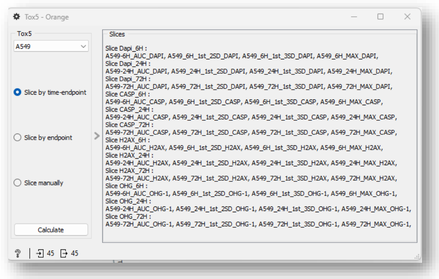
Read HTS data

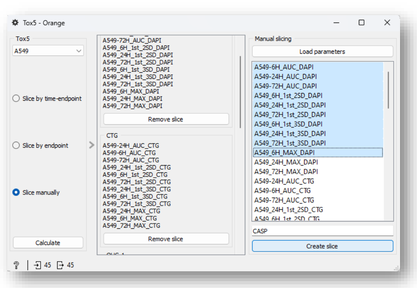
✓ Read HTS meta data from annotation file
✓ Read HTS data from csv or excel formats

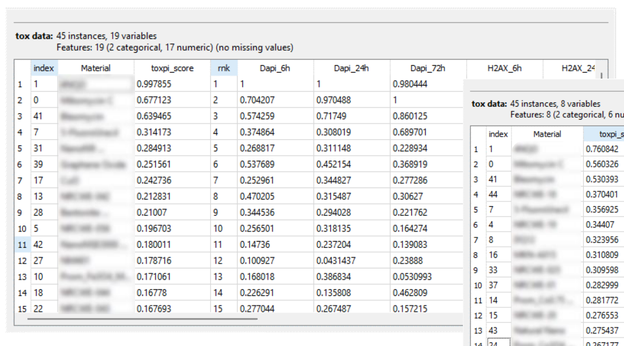


Tox5 calculation model

✓ Cell selection
✓ Automatic grouping by endpoint or by time-endpoint
✓ Manual creating specific groups

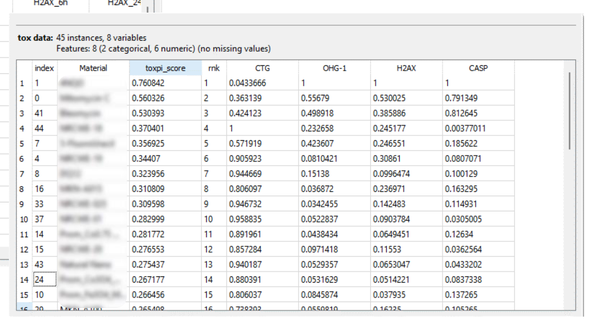






tox data: 45 instances, 19 variables
Features: 19 (2 categorical, 17 numeric) (no missing values)

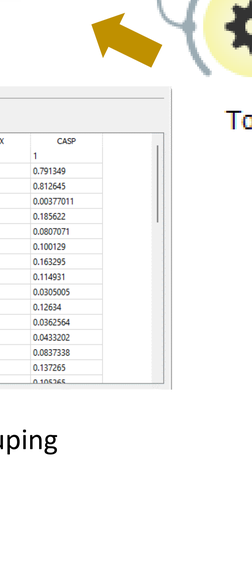
	index	Material	toxpi_score	rnk	Dapi_6h	Dapi_24h	Dapi_72h	H2AX_6h	H2AX_24
1	1	[redacted]	0.997855	1	1	1	0.980444		
2	0	Mitomycin C	0.677123	2	0.704207	0.970488	1		
3	41	Bleomycin	0.639465	3	0.574259	0.71749	0.860125		
4	7	5-Fluorouracil	0.314173	4	0.374864	0.308019	0.689701		
5	31	Nanofill9	0.284913	5	0.268817	0.311148	0.228934		
6	39	Graphene Oxide	0.251561	6	0.537689	0.452154	0.368919		
7	17	CuO	0.242736	7	0.252961	0.344827	0.277286		
8	13	NRCWE-056	0.212831	8	0.470205	0.315487	0.30627		
9	28	Bentonite (Nanoclay)	0.21007	9	0.344536	0.294028	0.221762		
10	5	NRCWE-056	0.196703	10	0.256501	0.318135	0.164274		
11	42	NanoNRCWE-042	0.180011	11	0.14736	0.237204	0.139083		
12	27	NRWE-041	0.178716	12	0.100927	0.0431437	0.23888		
13	10	Prom_Fe3O4_M...	0.171061	13	0.168018	0.386834	0.0530993		
14	18	NRCWE-042	0.16778	14	0.226291	0.135808	0.462809		
15	22	NRCWE-041	0.167693	15	0.277044	0.267487	0.157215		



tox data: 45 instances, 8 variables
Features: 8 (2 categorical, 6 numeric) (no missing values)

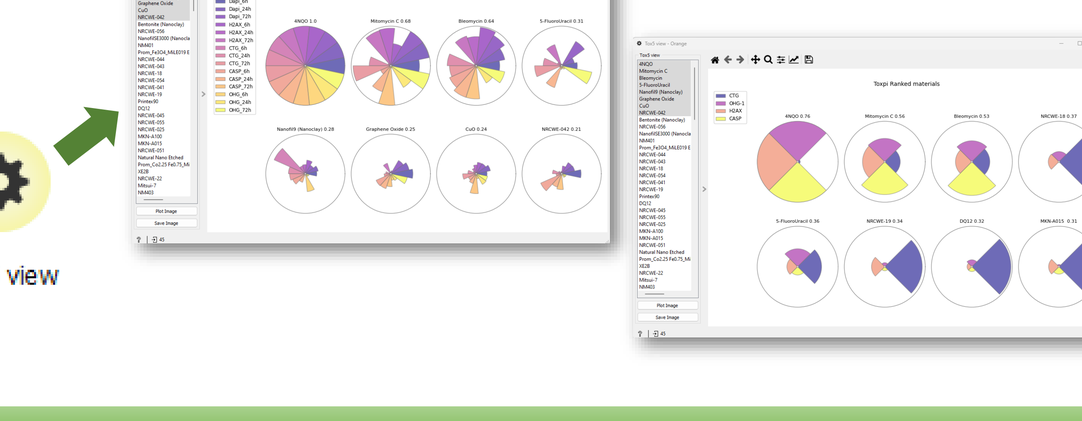
	index	Material	toxpi_score	rnk	CTG	OHG-1	H2AX	CASP
1	1	[redacted]	0.760842	1	0.0433666	1	1	1
2	0	Mitomycin C	0.560326	2	0.363139	0.55679	0.530025	0.791349
3	41	Bleomycin	0.530393	3	0.424123	0.498918	0.385886	0.812645
4	44	NRCWE-18	0.370401	4	1	0.232658	0.245177	0.00377011
5	7	5-Fluorouracil	0.356925	5	0.571919	0.423607	0.246551	0.185622
6	4	NRCWE-19	0.34407	6	0.905923	0.0810421	0.30861	0.0807071
7	8	DQ12	0.323956	7	0.944669	0.15138	0.0996474	0.100129
8	16	NRWE-037	0.310809	8	0.806097	0.036872	0.236971	0.163295
9	33	NRCWE-032	0.309598	9	0.946732	0.0342455	0.142483	0.114931
10	37	NRCWE-037	0.282999	10	0.958835	0.0522837	0.0903784	0.0305005
11	14	Prom_Co0.75...	0.281772	11	0.891961	0.0438434	0.0649451	0.12634
12	15	NRCWE-038	0.276553	12	0.857284	0.0971418	0.11553	0.0362564
13	43	Natural Nano	0.275437	13	0.940187	0.0529357	0.0653047	0.0433202
14	24	Prom_Co0.25...	0.267177	14	0.880391	0.0531629	0.0514221	0.0837338
15	10	Prom_Fe3O4_M...	0.266456	15	0.806037	0.0845874	0.037935	0.137265
16	20	MXN-A100	0.265408	16	0.728202	0.0550810	0.16225	0.105265

Table output with calculated TOX5 scores, ranks for different grouping



Toxpi preprocess

✓ Custom data normalization
✓ Calculating dose – response metrics



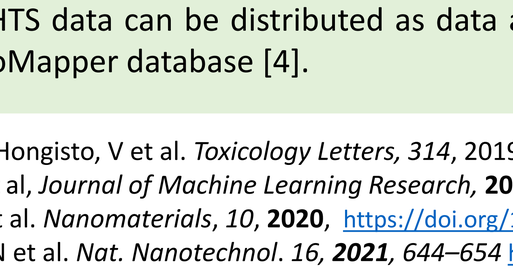
tox data: 45 instances, 241 variables
Features: 241 (1 categorical, 240 numeric) (20.6% missing values)

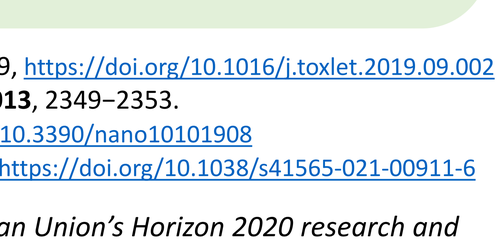
	material	A549-6H_AUC_DAPI	A549-24H_AUC_DAPI	A549-72H_AUC_DAPI	BEAS-2B-6H_AUC_DAPI	BS-2B-24H_AUC_DA
1	Mitomycin C	-0	122.905	308.365	-0	84.5813
2	4NQO	71.1447	173.685	262.075	158.043	276.574
3	NRCWE-055	31.887	78.7262	87.4904	18.2438	26.3497
4	Reduced ...	-0	13.8281	4.49606	-0	-0
5	NRCWE-19	-0	-0	-0	6.98796	-0
6	NRCWE-056	85.2529	116.004	115.073	118.993	114.356
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

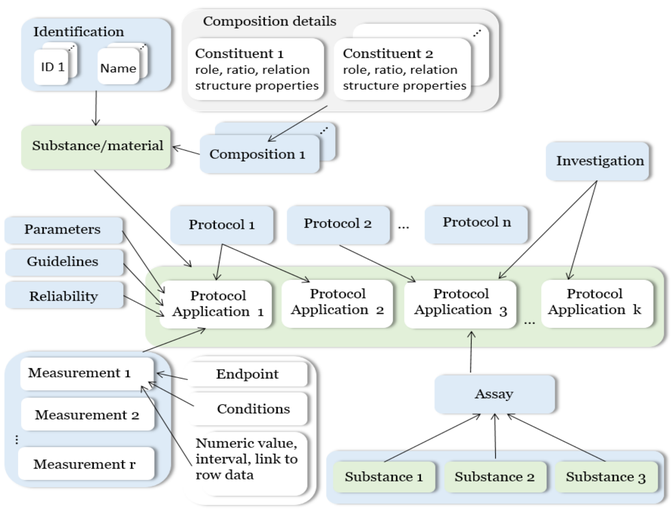
Tox5 pie view

✓ Pie view of ranks, and combined toxicity scores for each material
✓ User selectable set of materials for view
✓ Save image in convenient formats (e.g. png, jpeg)



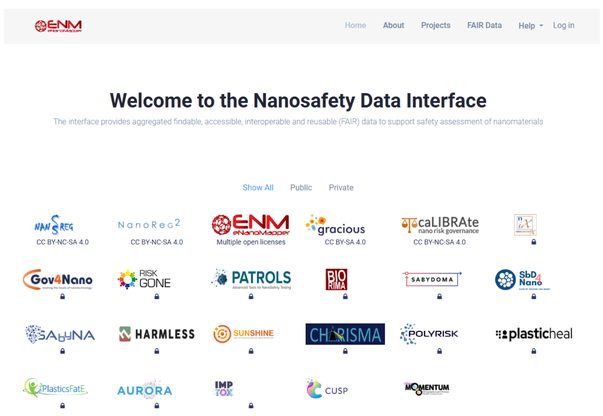


FAIR data in eNanoMapper



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].

eNanoMapper data model is flexible and 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].

Results

✓ 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**, https://doi.org/10.3390/nano10101908
4. Jeliazkova, N et al. *Nat. Nanotechnol. 16, **2021**, 644–654* https://doi.org/10.1038/s41565-021-00911-6

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