AUTOMATIC WORKFLOW FOR IN VITRO HIGH-THROUGHPUT SCREENING DATA FAIRIFICATION, PREPROCESSING AND SCORING: A CASE STUDY ON NANOMATERIALS

G. Tancheva[a,b], V. Hongisto[d], P. Nymark[c], K. Patyra[d], L. Iliev[a], N. Kochev[a,b], P. Kohonen[c], R. Grafström[c,d], N. Jeliazkova[a]

[a] *Ideaconsult, Bulgaria*

[b] *University of Plovdiv, Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, Bulgaria*

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

[d] *Misvik biology, Division of Toxicology Finland*

MOTIVATION

- Address safety challenges of new chemical substances and materials, including advanced nanomaterials.
- 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 high-throughput screening (HTS) biological data is used for efficient clustering, ranking, prioritization of NMs and read across.

CHALLENGES

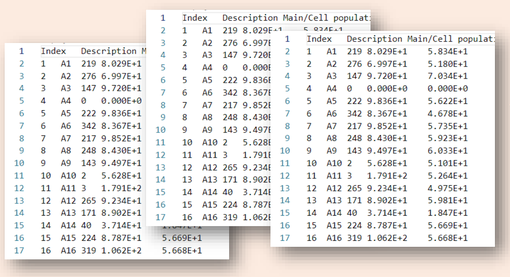
- Difficulty in consistently assessing and validating multiple agent effects simultaneously in HTS data.
- Challenges in linking large experimental datasets with descriptive metadata and harmonizing terminology.
- Ensuring data is machine-readable, findable, accessible, and reusable (FAIR principles) is complex.
- Traditional HTS results documentation approaches, such as using spreadsheets for data collecting and preprocessing are time-consuming and error prone.
- Integrating external tools like ToxPi [1] adds complexity, especially due to limited preprocessing capabilities and output options.

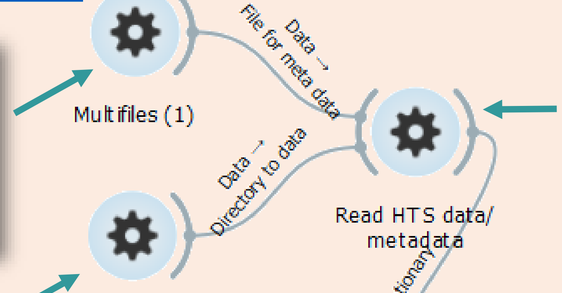
RESULTS

- ✓ **ToxFAIRy** [2] is a Python package that provides faster data preprocessing, scoring, hierarchical clustering, and data FAIRification. ToxFAIRy minimizes the possibility of errors through automation and improves on the original ToxPi software in terms of data visualization, flexibility, and easier data export.,
- ✓ **Orange3-ToxFAIRy [2]** is an Orange3 add-on that makes composition of complex workflows much easier for the users through a visual programming interface.
- ✓ **HTS_METADATA template** developed as part of the Template Wizard supports data annotation and harmonization in a reproducible manner
- ✓ 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 serialized in NeXus format that can be integrated into the eNanoMapper database [4].

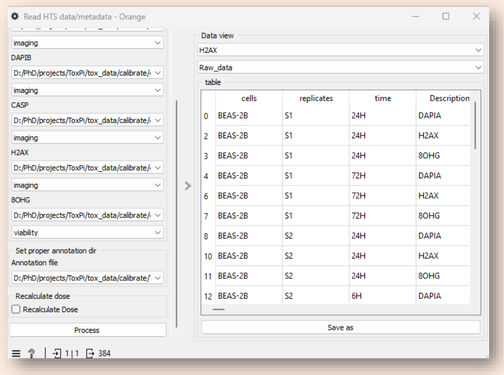
Case study on NANOMATERIALS

The example raw data and filled out HTS_METADATA templates are available in https://doi.org/10.5281/zenodo.13683162

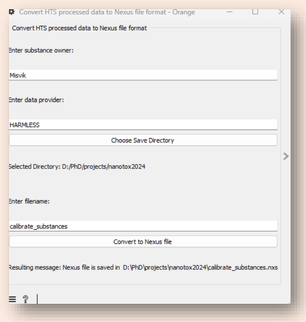


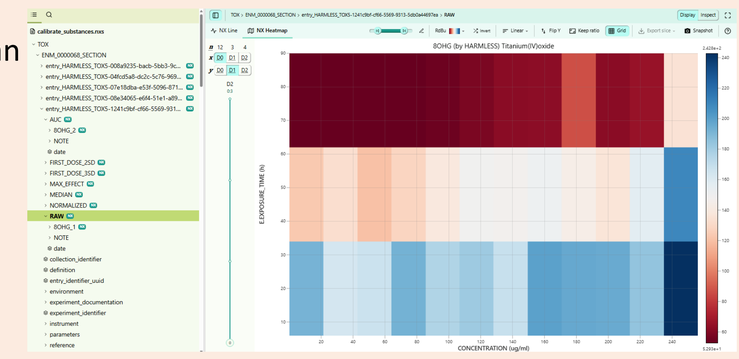


Read and annotate HTS data with option for dose recalculation



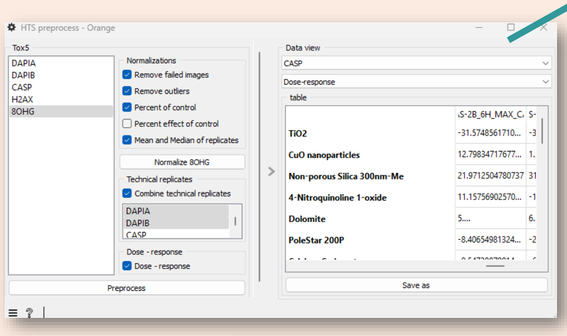
Make the processed data FAIR by converting to an eNanoMapper data model and serialize in NeXus format.



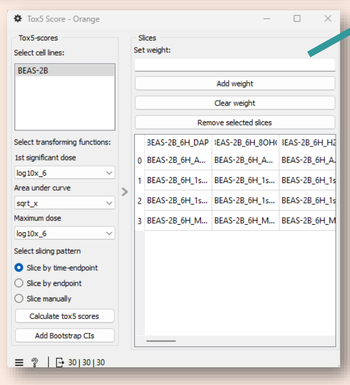


Visualize NeXus file from: https://myhdf5.hdfgroup.org/

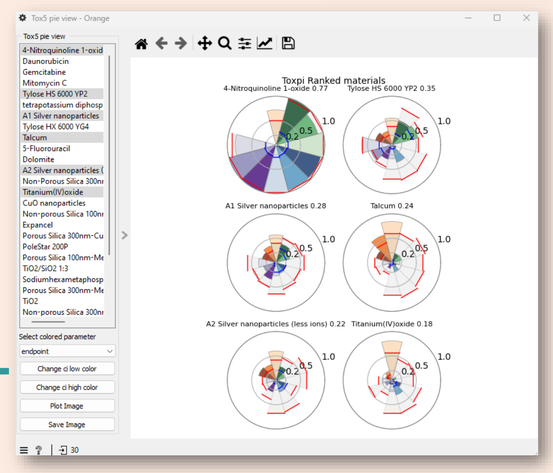
ToxFAIRy [2] collects and annotates raw HTS data in a harmonized way, normalizes and calculates dose-response parameters, and provides data FAIRification and serialization in NeXus format. It integrates the ToxpiR library and strictly follows the original Tox5-score [5] approach. ToxFAIRy can be used as a library itself, but we also provide an Orange [6] add-on, Orange3-ToxFAIRy [2], with user-friendly widgets for fine-tuning of the data processing.



Normalize and calculate dose-response metrics for each endpoint separately.

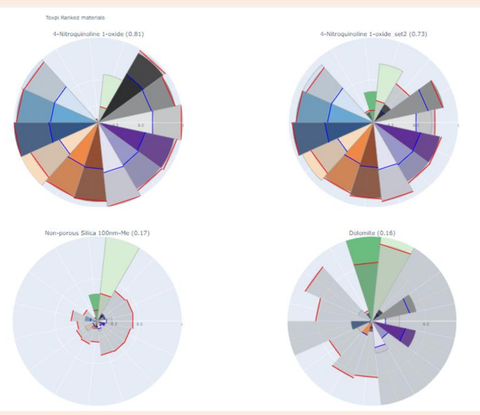


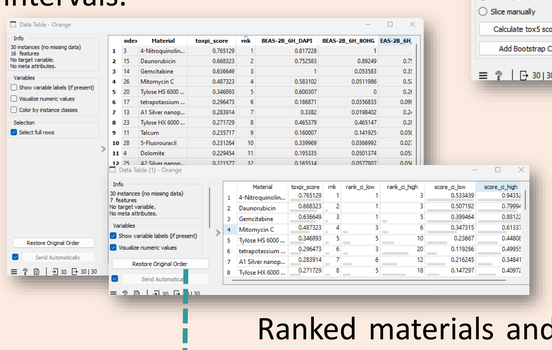
Create slices automatically or manually for selected cell lines. Transform data with chosen functions, add specific weight for selected metrics and calculate scores with Bootstrap confidence intervals.



Pie view of toxicity scores of selected materials with upper and lower bounds of confidence intervals.

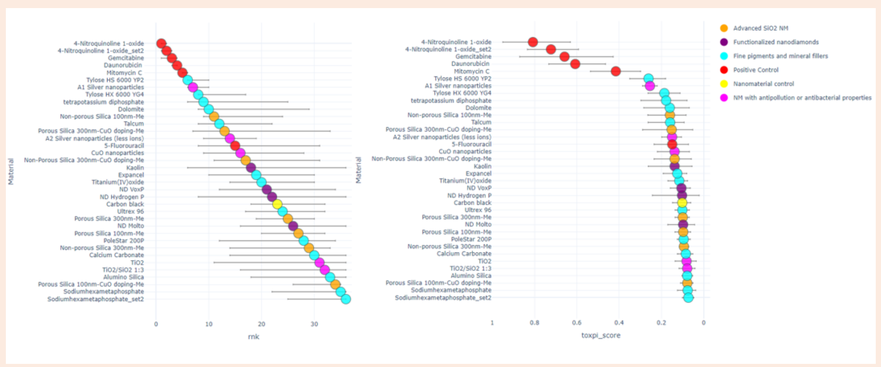
The pie chart is available as an interactive HTML graphic, featuring slice markers that display additional information on hover, along with zoom functionality for enhanced exploration.





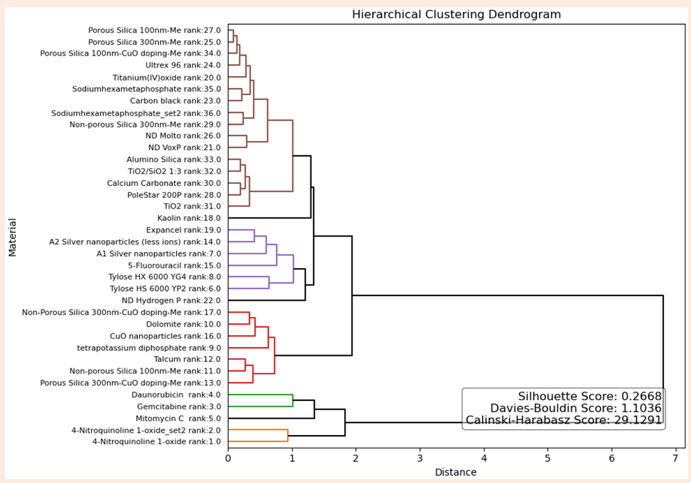
Resulting tables with Tox5-scores, ranks, specific slices scores and confidence intervals

Ranked materials and controls with bootstrap confidence interval from resulting table

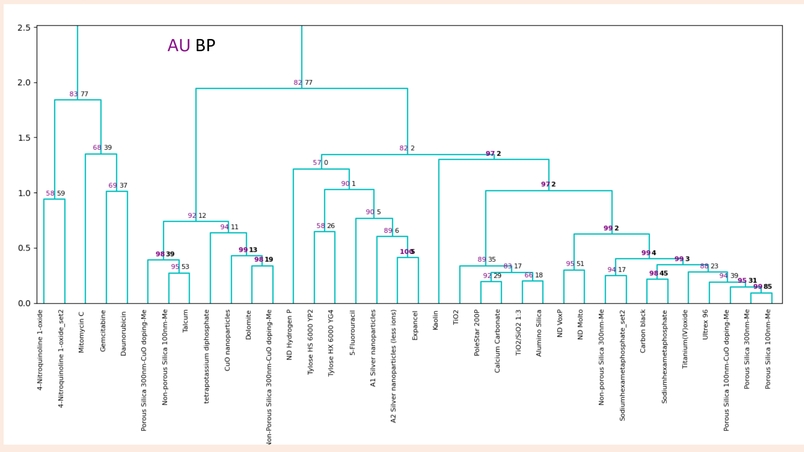


Hierarchical clustering methods integrated in ToxFAIRy

Hierarchical clustering, utilizing euclidean distance as the metric and clustering method – ward and optimal number of clusters by elbow method. Automatically calculated cluster significance metrics Silhouette, Davies-Bouldin and Calinski-Harabasz scores.



Multiscale bootstrap resampling to the hierarchical clustering, by reintegrated pvclust python package. Approximately Unbiased p-value (AU) and Bootstrap Probability (BP) report the significance of each cluster where clusters > 95% are considered significant and strongly supported by the data.



1. Marvel, S.W., et al. BMC Bioinformatics 19, 80, 2018 https://doi.org/10.1186/s12859-018-2089-2
2. https://github.com/ideaconsult/orange3-toxfairy
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
5. Nymark, P; Hongisto, V et al. *Toxicology Letters, 314*, 2019, https://doi.org/10.1016/j.toxlet.2019.09.002
6. Demsar, J et al, *Journal of Machine Learning Research,* 2013, 2349−2353.

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