Automatic workflow for in vitro high-throughput screening data FAIRification, preprocessing and scoring

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

HARMLESS
https://www.harmless-project.eu/

BACKGROUND MOTIVATION AND OBJECTIVE

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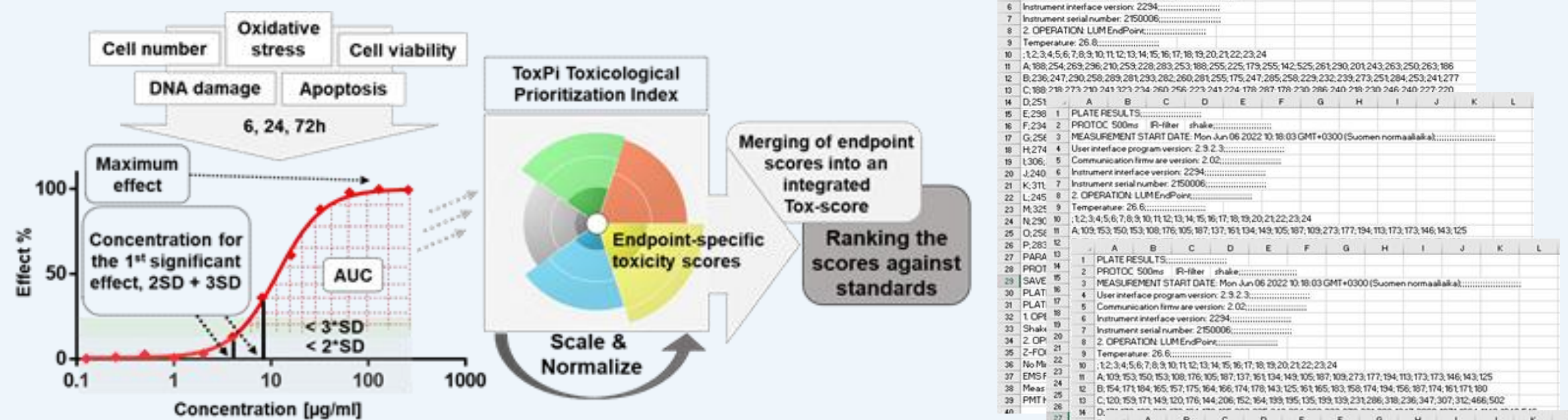
Tox5-score [1] in vitro toxicity scoring and ranking concept:

- 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
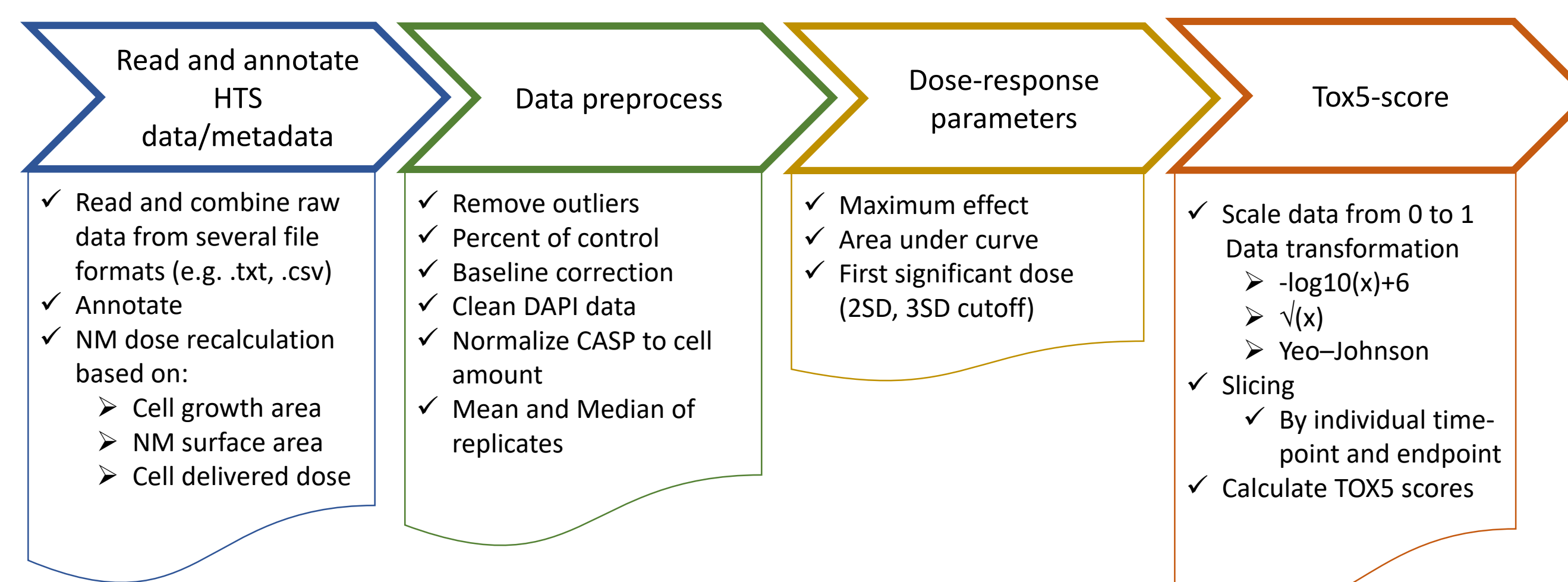
HTS generates vast amounts of data, which are difficult to process and analyze. Toxicological Priority Index [2] is a computational tool used in toxicology and environmental health sciences to assess and prioritize the potential health risks associated with exposure to multiple chemical stressors.

Challenges when using the **ToxPi** software

➢ requires time-consuming manual processing
➢ difficult to scale for larger NM datasets
➢ prone to errors
➢ challenging to implement FAIR principles

METHODS

A new Python module **TOX5** (tentative name) [3] has been developed for the collection and annotation of raw data, followed by normalization and the calculation of dose-response metrics. This module relies on the ToxpiR library as a key dependency and strictly follows the original Tox5-score approach. The **TOX5** module can be used as a library itself, but we also provide an Orange [4] add-on **Tox5-scores** [3] with user-friendly widgets for fine-tuning of the data processing.
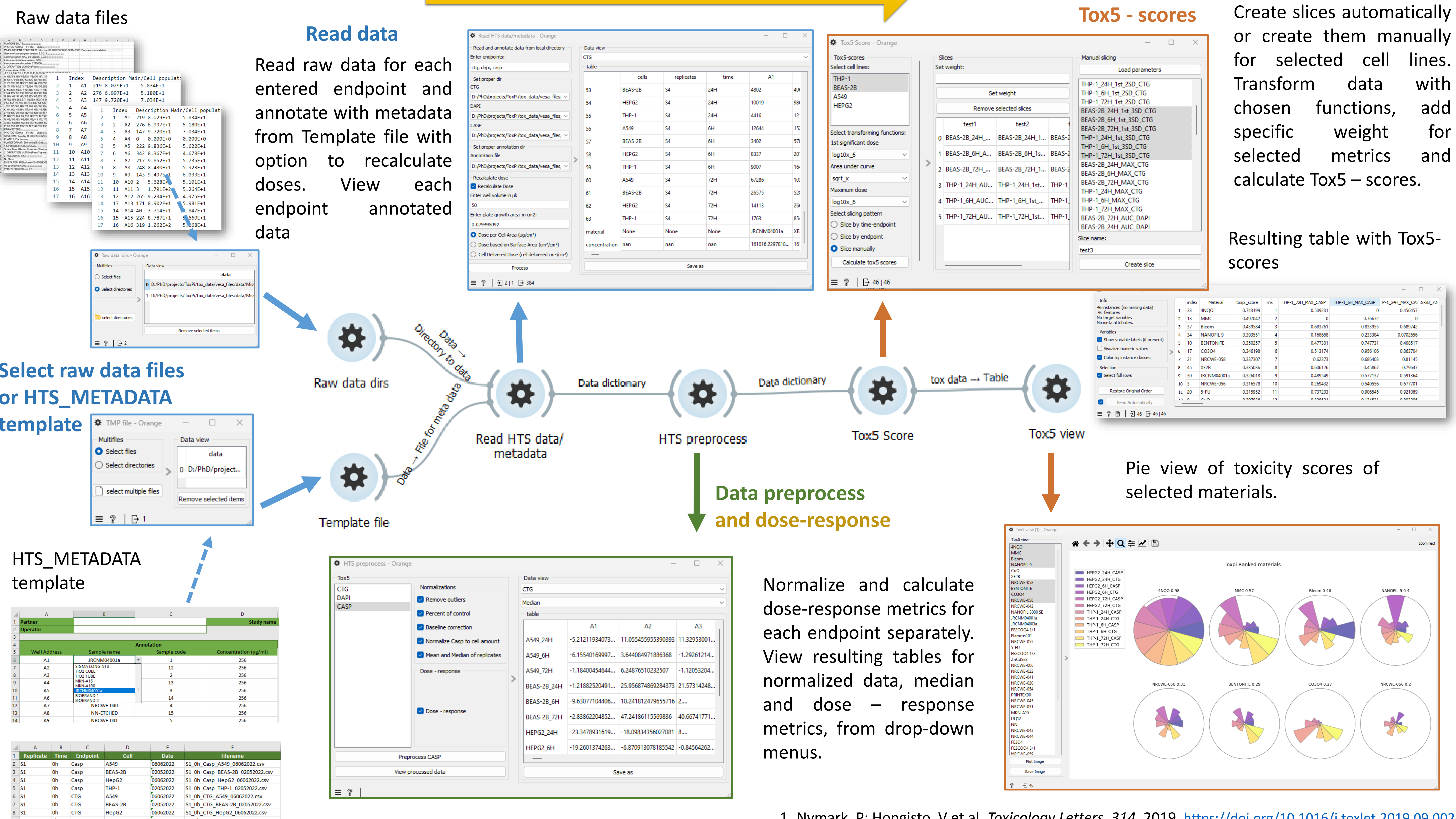
General steps in workflow for HTS data preprocessing and toxicity scoring

Read and annotate HTS data/metadata	Data preprocess	Dose-response parameters	Tox5-score
✓ Read and combine raw data from several file formats (e.g. .txt, .csv) ✓ Annotate ✓ NM dose recalculation based on: ➢ Cell growth area ➢ NM surface area ➢ Cell delivered dose	✓ Remove outliers ✓ Percent of control ✓ Baseline correction ✓ Clean DAPI data ✓ Normalize CASP to cell amount ✓ Mean and Median of replicates	✓ Maximum effect ✓ Area under curve ✓ First significant dose (2SD, 3SD cutoff)	✓ Scale data from 0 to 1 Data transformation ➢ -log10(x)+6 ➢ √(x) ➢ Yeo–Johnson ✓ Slicing ➢ By individual time-point and endpoint ✓ Calculate TOX5 scores

Results

✓ The new Python module **TOX5** enables faster data preprocessing calculations and minimizes the possibility of errors through automation.
✓ The **TOX5** Python module and the **Tox5-scores** Orange workflow are extending the eNanoMapper FAIRification workflow [5] 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 the eNanoMapper database [6].

Tox5-scores automatic workflow

Raw data files

Read data

Read raw data for each entered endpoint and annotate with metadata from Template file with option to recalculate doses. View each endpoint annotated data

Tox5 - scores

Create slices automatically or create them manually for selected cell lines. Transform data with chosen functions, add specific weight for selected metrics and calculate Tox5 – scores.

Resulting table with Tox5-scores

Select raw data files or HTS_METADATA template

Raw data dirs

Template file

HTS_METADATA template

Read HTS data/ metadata → HTS preprocess → Tox5 Score → Tox5 view

Data → Directory to data
Data → File for meta data
Data dictionary → Data dictionary → tox data → Table

Data preprocess and dose-response

Normalize and calculate dose-response metrics for each endpoint separately. View resulting tables for normalized data, median and dose – response metrics, from drop-down menus.

Pie view of toxicity scores of selected materials.

1. Nymark, P; Hongisto, V et al. *Toxicology Letters, 314*, 2019, https://doi.org/10.1016/j.toxlet.2019.09.002
2. Marvel, S.W., et al. BMC Bioinformatics 19, 80, 2018 https://doi.org/10.1186/s12859-018-2089-2
3. https://github.com/ideaconsult/orange-tox5
4. Demsar, J et al, *Journal of Machine Learning Research*, **2013**, 2349–2353.
5. Kochev, N et al. *Nanomaterials*, 10, **2020**, https://doi.org/10.3390/nano10101908
6. Jeliazkova, N et al. *Nat. Nanotechnol*. 16, **2021**, 644–654 https://doi.org/10.1038/s41565-021-00911-6