# Are We Doing Non-targeted Analysis Right? A Progress Report from the **Benchmarking and Publications for Non-Targeted Analysis Working Group**

Elin M. Ulrich<sup>1</sup> (ulrich.elin@epa.gov), Benjamin J. Place<sup>2</sup> (benjamin.place@nist.gov)

1. U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC 2. National Institute of Standards and Technology, Material Measurement Laboratory, Chemical Sciences Division, Organic Chemical Measurement Science Group, Gaithersburg, MD

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency or National Institute of Standards and Technology.

## **Definitions in Progress**

## Examples of Working Group Progress

Group 1: Identification

up 6: Experimental Design Terms

detected f

Components of a Data Analysis Method:

Raw data processing: Includes centroiding, smoothing, thresholding.

Peak picking: Selecting unidentified peaks/mass features/mz-rt pairs

Library Searching: Using a library/database to match spectra/precursor

masses between experimental and known info, to include match score.

In silico fragmentation generation: Generating theoretical spectrum to

Discussion: Define all 11? All applicable to NTA?

was produced and its intended use - don't let the

What is used depends strongly on the analytical task

New strategy: More important to detail HOW the blank

Mass spectral interpretation: Annotation of m/z fragments to "build"

reader interpret for themselves.

"User Expertise" : Need to report exactly how this was applied.

Revised definition:

Identification/Annotation: attribution of

ture within a sample with

re- is there a better term?

a chemical identity (Identification) or

chemical formula (Annotation) of a

an associated confidence

#### Category more common for environmental and other applications, particularly for suspect Experimenta screening and non-targeted analysis (NTA). While fields like metabolomics have developed mature methodologies and quality control practices, NTA of the exposome is still experiencing a steep learning curve and growing pains. The current state of NTA has been described as "the wild west," where each research group approaches the technique in their own way, with little overlap, consistency, or harmonization. In a 2018 EPA NTA workshop, a discussion resulted in the formation of a working group called "Benchmarking and Publications for Non-Targeted Analysis" or BP4NTA. The group has a near term goal of publishing a white paper describing terms, definitions, recommendations, and best practices surrounding NTA studies. Topics of particular importance and relevance will include recommendations on how to characterize an NTA method's performance and minimum reporting information for publications to improve transparency and General reproducibility. If NTA exposome data is to be used to support further scientific endeavors and/or for regulatory purposes, practitioners should use sound, defensible, and commonly accepted techniques for data collection, analysis, and reporting. Therefore, BP4NTA and other groups like it in related fields are critical for coming to a scientific-community consensus to make an impact on environmental and health policies Interpretatio About BP4NTA

Performance

Statistics

QA/QC

Repeatability

Reproducibility

61 members: 26 government, 20 industry, 15 academia

ited States **Environmental Protection** 

al Institute of

The broad availability of high-resolution mass spectrometers has made their use

ds and Tech

Agency

WP123

Abstract

- > Monthly conference calls or in-person meetings
- > Collaborative documentation/discussions via Google Drive
- 9 working groups for definitions (see Table)

### Short term goals:

#### Publish a white paper containing: NTA terms and their definitions

- Calculations of performance metrics:
- · Reporting recommendations to promote transparency/reproducibility; Scientific best practices.
- 2. Build consensus with like-minded organizations (e.g., NORMAN).
- Work with journal editors and NTA researchers to establish guidelines NTA

reporting for methods and performance evaluations.

Long term goal (~10 yrs): Move the field of NTA toward proficiency testing, using a mechanism like ASTM/ISO Guidance on Performance and Data Reporting Requirements. Define proficiency levels for SSA, NTA, expert, competent, etc.

> JOIN US! Leave your card or contact the poster authors Ben or Elin.

/	Terms and Sub-terms to Define	Working Group Assignments	Group 1: Identification
l Design	Blank Laboratory- Matrix- Solvent- Replicate -Data analysis -Extraction -Injection -Sample Laboratory standard Spike	Group 6: Blank + Matrix blank + Spike + Laboratory standard + Replicate	BPANTA Initial Definition:   Identify (specific to the stereoisomer form, at minutum, within a sample with an associated confidence; confidence level must be assigned for all identifications. Single chemical identity, could be identified to the formula level must be a pretty strong language Image: Comparison of the identifications. Single chemical identity, could be identified to the formula level must be a pretty strong language Image: Comparison of the identifications. Single chemical identity, could be identified to the formula level must be a pretty strong language Image: Comparison of the identified to the formula level must be a pretty strong language Image: Comparison of the identified to the formula level must be a pretty strong language Image: Comparison of the identified to the formula level must be a pretty strong language   BP4NTA Initial Definition: Raw data processing: Includes centry language Components of a Definition: Comparison of the identified to the formula level must be a pretty strong language   Mata analysis method is the treatment of the fare NTA data (with no processing) that (needs completion) Image: Compare with empirical spectrum. Mass spectral interpretation: Annotation: Types of blanks; Ambient, Calibration, Edupinent, Field, Filter, Fortified method, Matrix, Method, Preservation, Reagent, Trip. http://www.chromatographyonline.com/vital-role-blanks-sample-preparation?pageID=1 Discussion: Define a was produced and i eader interpret for the identified of the must be added to the interpret for the identified of
	Methods Data analysis- Instrumental- Sample preparation- Library/database Non-targeted analysis Suspect screening analysis	Group 2: Non-targeted analysis + Suspect screening Group 4: Data analysis method Group 5: Library/database Group 7: Instrumental method Group 9: Sample preparation method	
n of Data	Identification Confidence of identification	Group 1: Identification + Confidence of identification	
3	True positive (TP), TP rate/ratio True negative (TN), TN rate/ratio False negative (FN), FN rate/ratio False positive (FP), FP rate/ratio Unintended positive (UP), UP rate/ratio False discovery rate (FDR) Negative predictive value Accuracy Performance Specificity Precision Sensitivity, recall Area under precision-recall (AUPR) curve F1 Score	Group 3: False positive + True positive + True negative	
	Accuracy Precision	Group 8: Performance + Accuracy + Reproducibility + Precision	> Continue to refine definitions/equations, build/refine "components of" lists

Reproducibility + Precision

**HOW WOULD YOU DEFINE THESE TERMS? LEAVE A POST-IT WITH YOUR IDEAS!** 

ons, build/refine "components of" lists, build reference library in BP4NTA Discussion/refinement/(dis)agreement within BP4NTA Develop a list of questions/topics for discussion within broader NTA community. Is consensus possible? Develop outline and draft of white paper. Publish white paper Communicate. Communicate! COMMUNICATE and obtain buy in from the broader NTA community. Put ideas into practice as practitioners, editors, reviewers, mentors, etc.