



Quality assurance and quality control workflow for the non-targeted analysis of de facto water reuse

Presented by Nick Sayresmith (ORISE Postdoc at EPA)

Laura D. Brunelle, Angela L. Batt, Alex Chao, E. Tyler Carr, Jon R. Sobus

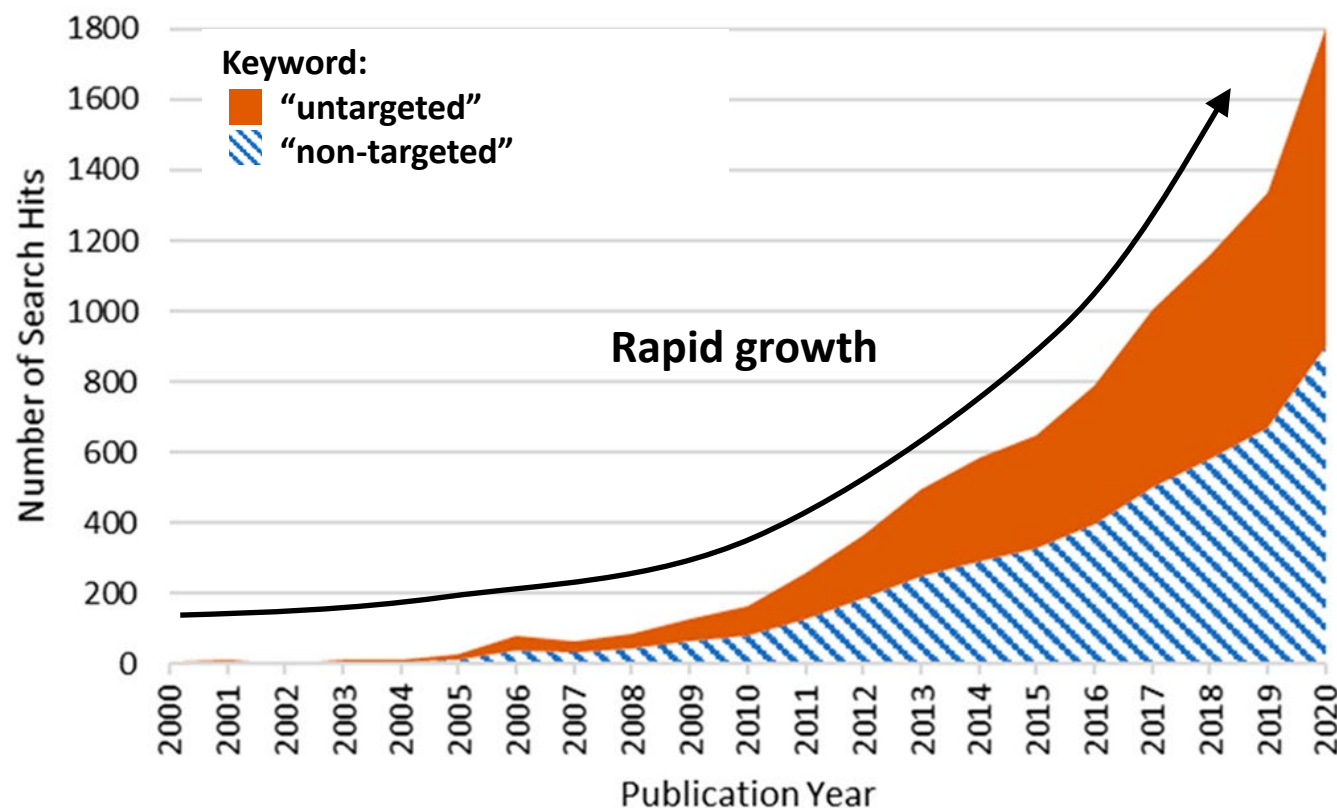
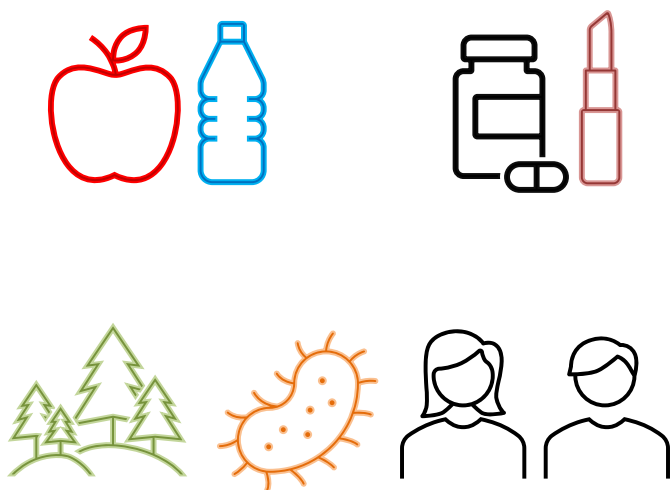
De Facto Water Reuse Study PI's – Dave A. Alvarez, Ed T. Furlong, Susan T. Glassmeyer, Dana W. Kolpin, Marc A. Mills

WebApp Contributors - T. Cathey, J.P. McCord, J.M. Minucci, S.T. Purucker, A. Rashid, D. Smith, A.J. Williams

Non-Targeted Analysis (NTA): Definition and State-of-the-Science

- NTA – method for characterizing the chemical composition of a sample without *a priori* knowledge of the sample's chemical content
 - Whereas targeted analysis relies on pre-selected standards

Wide applicability and broad interest



Non-Targeted Analysis (NTA): State-of-the-Science

- Still relatively nascent methodology
 - Limited guidance and standardization for study design and reporting (sample prep, data acquisition, **QA/QC metrics**)
 - NTA Study Reporting Tool (SRT) helping to address challenges

A FRAMEWORK FOR CONSISTENT PEER REVIEW

NTA Study Chronology

Section	Category	Sub-Category	Example Information to Report	Score	Rationale
Methods	Study Design	Objectives & Scope	• Study goals, hypotheses, scope • Expected chemical coverage	1	Space for reviewer to explain assigned score (i.e., typical peer review rationale)
		Sample Info & Preparation	• Sampling collection, processing • Description, intended use of blanks	2	
		QC Spikes & Samples	• Description of spikes/controls	2	
	Data Acquisition	Analytical Sequence	• Sample run order, analytical batch(es)	NA	
		Chromatography	3-4 examples of representative information to report for each of the 13 sub-categories. - Not an exhaustive list – intended to guide researcher/reviewer and relies on expertise/discretion.	0	
		Mass Spectrometry		1	
	Data Processing & Analysis	Data Processing		2	
		Statistical & Chemometric Analysis		3	
		Annotation & Identification			
	Results	Data Outputs	Statistical & Chemometric Outputs	• Basic statistical outputs & results of chemometric analyses • Results/plots, new statistical metrics, algorithms, etc.	
ID & Confidence Levels			• Reported IDs and confidence levels & supporting data • (Semi)-quant data, exported MS/MS spectra	3	
QA/QC Metrics		Data Acquisition QA/QC	• Method impacts on observable chemical space • Accuracy & precision of chromatography, mass error, abundance	1	
		Data Processing & Analysis QA/QC	• Method impacts on observable chemical space • Performance measures for accuracy, reproducibility of results	0	

Enables rigorous evaluation of reporting quality in NTA studies

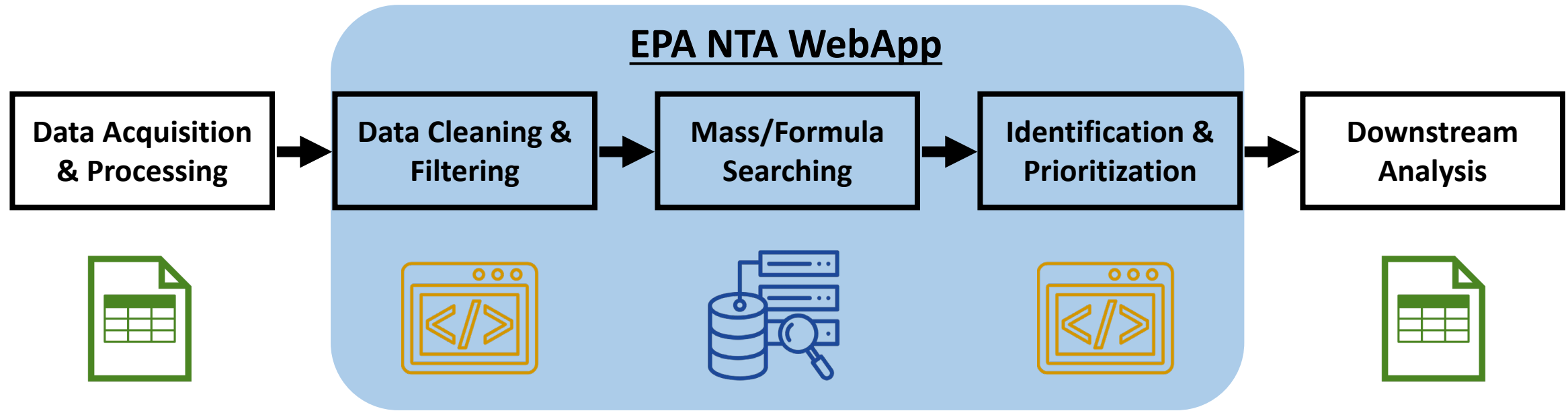
Enables rigorous evaluation of reporting quality in NTA studies



<https://nontargetedanalysis.org/>

<https://doi.org/10.1021/acs.analchem.1c02621>

Typical NTA Data Workflow, Common Issues, and NTA WebApp



Common Issues:

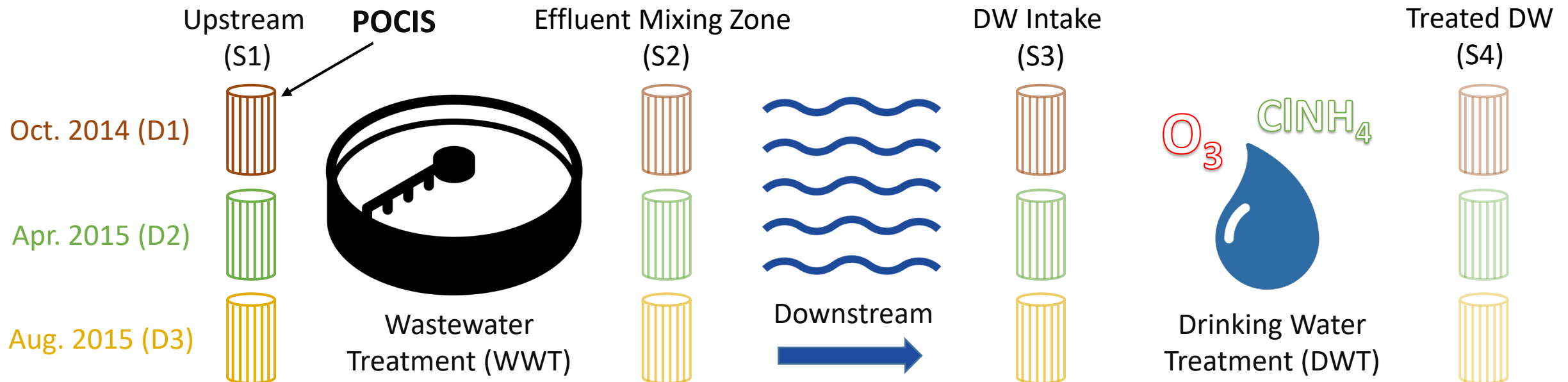
- Individualized methodologies and/or custom code lead to
 - High variability between practitioners
 - Opaque processes
 - Difficult to reproduce and communicate
 - User burden and slow adoption

Goals of WebApp:

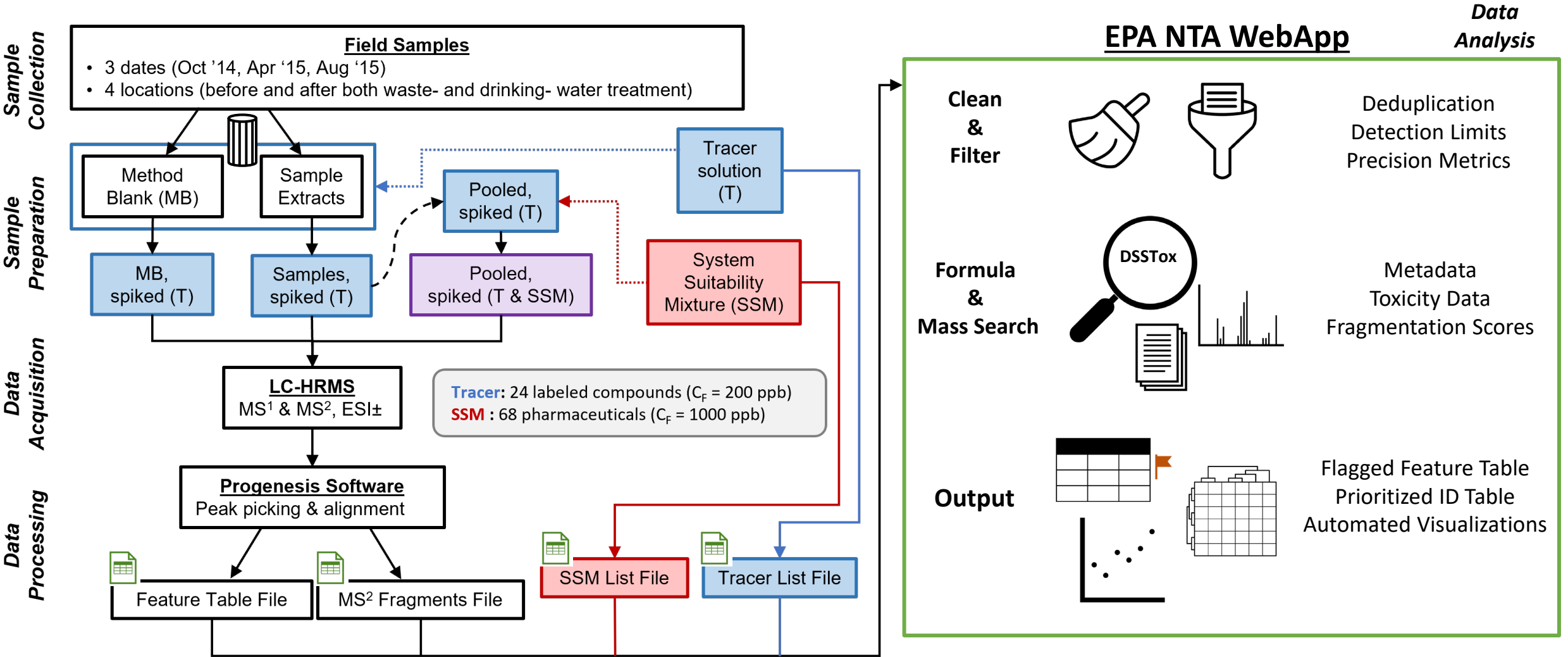
- **Standardize** by using statistically rigorous methodology and robust code
- **Elucidate** via guides and tool tips
- **Streamline** while maintaining user control

De Facto Water Reuse Study Summary

- **Goal** – examine the exposure and fate of contaminants of emerging concern (CECs) from before wastewater treatment to after drinking water treatment
- **Subgoal** – use passive sampling and NTA to broadly identify CECs
 - **POCIS** – polar organic chemical integrative samplers



Study Design and WebApp



NTA: non-targeted analysis of MS data (beta)

[Contact Us](#)

Tools

MS1 Tool

Run MS1 Tool

MS1 Tool Algorithms

MS1 Tool QA/QC

MS1 Tool References

MS2 CFMID Tool

Documentation

Source Code

Run NTA MS1 Tool

Easy file inputs

Input	Value
Project name:	Example nta
Positive MPP file (csv):	Choose File No file chosen
Negative MPP file (csv):	Choose File No file chosen
Adduct mass accuracy units:	ppm
Adduct mass accuracy:	10
Adduct retention time accuracy (mins):	0.05
Tracer file (csv; optional):	Choose File No file chosen
Tracer mass accuracy units:	ppm
Tracer mass accuracy:	5
Tracer retention time accuracy (mins):	0.1
Min sample:blank cutoff:	3
Min replicate hits:	<input type="range" value="2"/>
Max replicate CV:	0.8
Parent ion mass accuracy (ppm):	<input type="range" value="5"/>
Discard features below this retention time (mins):	0.0
Search dashboard by:	mass
Save top result only?	no
DSSTox search batch size (debugging):	150

Defaults

Clear

Save Metadata?

Submit

User-defined parameters

Screenshots of NTA WebApp

NTA: non-targeted analysis of MS data (beta)

[Contact Us](#)

Tools

MS1 Tool

MS2 CFMID Tool

Run MS2 CFMID Tool

MS2 CFMID Tool Algorithms

MS2 CFMID Tool QA/QC

MS2 CFMID Tool References

Merge MS1 and MS2 Data

Documentation

Source Code

Run MS2 CFMID Tool

Input	Value
Job ID:	DVE6FJR0
Neg Mode Data:	Choose File No file chosen
Pos Mode Data:	Choose File No file chosen
Project name:	Example nta
Precursor mass accuracy (ppm):	10
Fragment mass accuracy (Da):	0.02
Fragment Classifications (csv):	Choose File No file chosen

Clear

Defaults

Submit

Tracer Results Table

Chemical Name	DTXSID	Ionization Mode	Mass Error (ppm)	RT Drift (min)	Precision (max %CV)	Detection Frequency (%)
User-defined settings			< 10	< 0.5	< 50	100
acetaminophen-d3	DTXSID50480414	ESI+	3.20	0.06	12	100
albuterol-d9	DTXSID10675541	ESI+	2.52	0.00	20	100
amitriptyline-d6	DTXSID501349824	ESI+	2.14	0.09	18	100
amlodipine-d4	DTXSID50661983	ESI+	8.59	0.09	35	97.8
atenolol-d7	DTXSID101027977	ESI+	2.59	0.02	15	100
Caffeine-13C3	DTXSID20437172	ESI+	2.45	0.02	17	100
carbamazepine-d8	DTXSID401349821	ESI+	0.79	0.12	22	100
diltiazem-d3	DTXSID801016193	ESI+	1.61	0.05	15	100
fluoxetine-d5	DTXSID50661983	ESI+	0.41	0.10	18	100
glipizide-d11	DTXSID601349827	ESI+	4.51	0.13	20	100
metoprolol-d7	DTXSID30648858	ESI+	2.92	0.07	16	100
norethindrone-d6	DTXSID401349857	ESI+	0.48	0.11	14	100
paroxetine-d4	DTXSID101349822	ESI+	1.58	0.05	32	100
sertraline-d3	DTXSID201349825	ESI+	0.35	0.08	23	100
sulfamethoxazole-d4	DTXSID101016780	ESI+	3.68	0.09	15	100
triamterene-d5	DTXSID701349820	ESI+	1.25	0.05	29	100
trimethoprim-d9	DTXSID10662219	ESI+	2.19	0.04	15	100
verapamil-d6	DTXSID801349823	ESI+	0.39	0.08	28	100
warfarin-d5	DTXSID801016155	ESI+	1.62	0.17	19	100
gemfibrozil-d6	DTXSID601028063	ESI-	0.29	0.16	12	100
glipizide-d11	DTXSID601349827	ESI-	1.19	0.04	22	100
hydrochlorothiazide-13C,d2	DTXSID00662001	ESI-	0.92	0.06	8	100
ibuprofen-d3	DTXSID00481299	ESI-	1.01	0.15	12	100
sucralose-d6	DTXSID301339960	ESI-	1.62	0.01	12	100
warfarin-d5	DTXSID801016155	ESI-	0.13	0.03	11	100

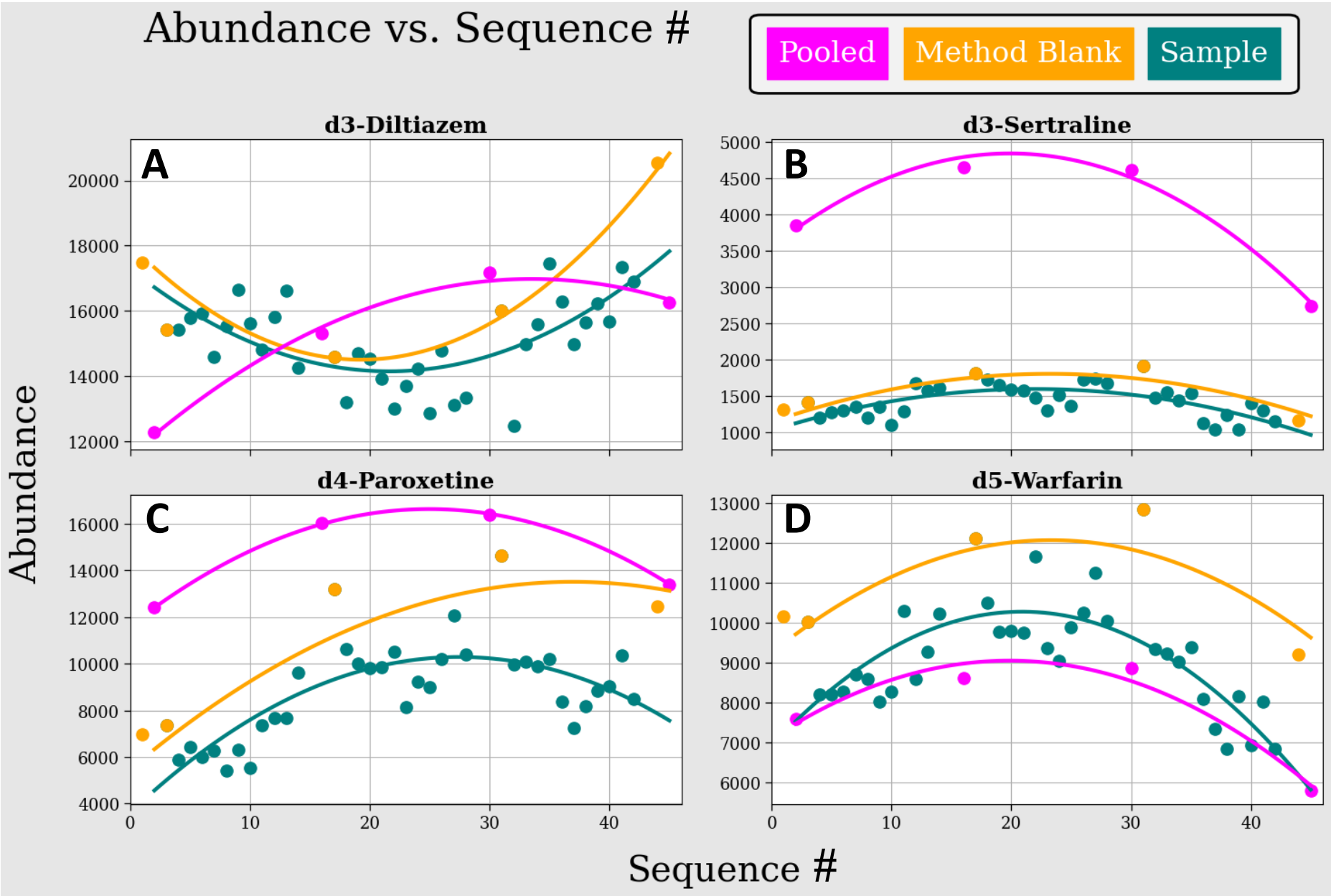
- Crucial first step for ensuring data integrity
- Provides hyperlinks to DSSTox database
- Summary statistics for each ionization mode
- Out-of-spec statistics prompt follow-up action
 - Doesn't necessarily invalidate whole study

*most extreme

Run Sequence Plot Demonstrating Sequence and Matrix Effects (ESI+)

No effect
(4)

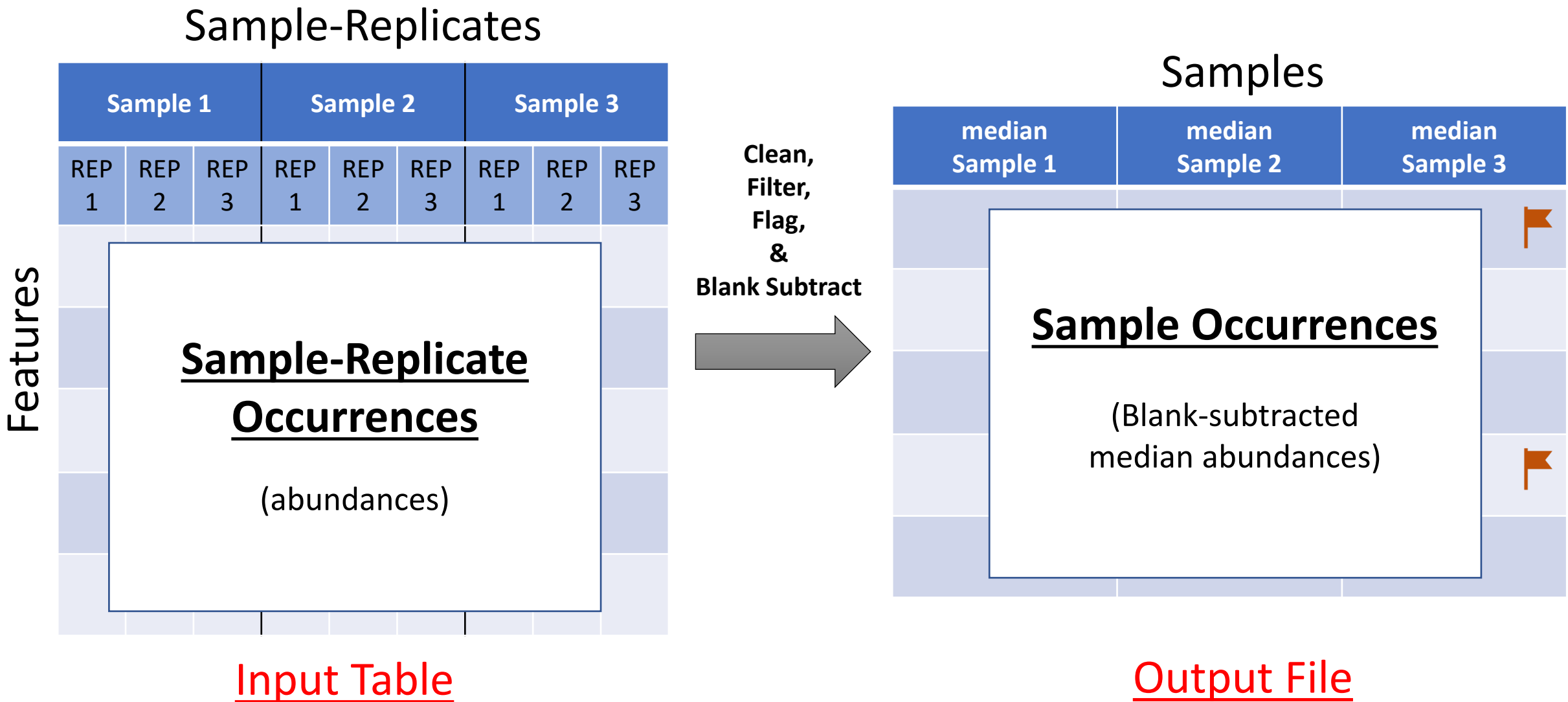
Miscellaneous
(1)



Enhancement
(5)

Suppression
(9)

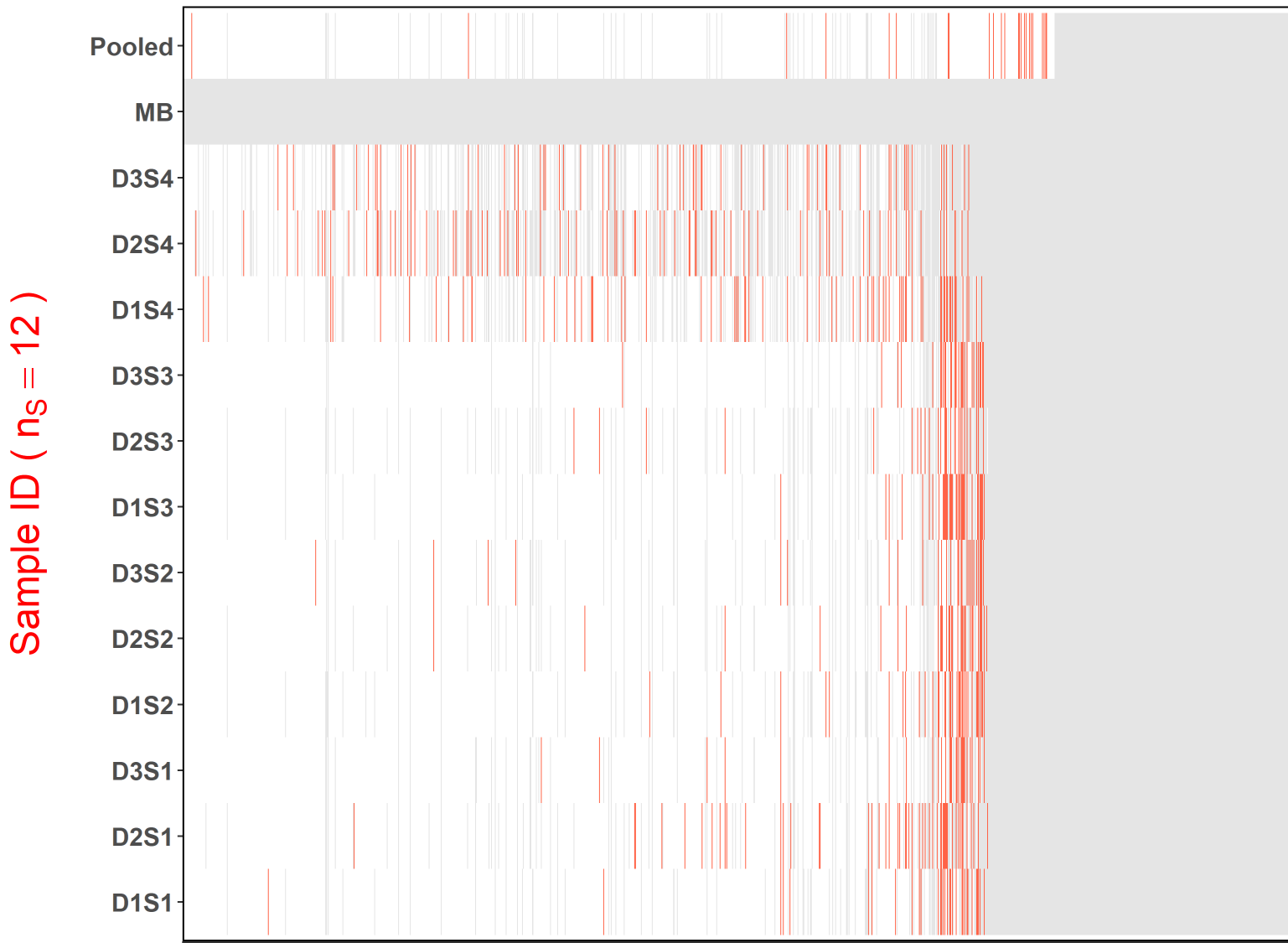
Nomenclature – features, samples, occurrences



CV Heat Map: A Visualization of Flagged, Non-Detected, and Passing Occurrences

ESI+, post-blank subtraction ($n_{Occ} = 15540$)

Sorted by median sample abundance (highest [left] to lowest [right])



Occurrence Flag (samples only)

- ≤ 0.5 (8594, 55%), no flags
- > 0.5 (460, 3%), CV flag
- non-detect (6486, 42%)

Includes:

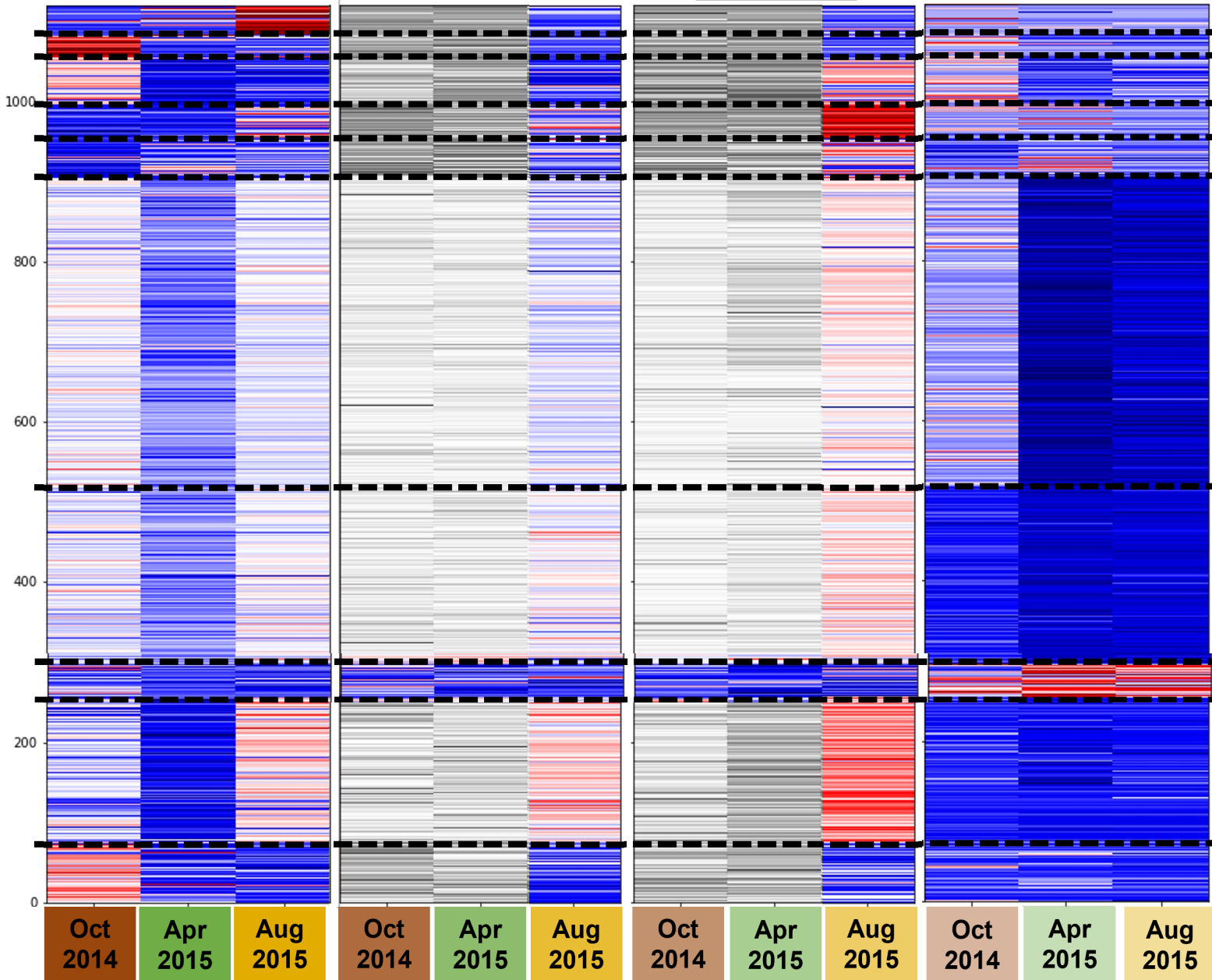
- missing all replicates
- $<$ replicate threshold
- $<$ MDL

Upstream

Effluent Mixing Zone

DW Intake

Treated DW



Cluster Heat Map (k-means)

- Differences across time?
 - Seasonality and stream flow conditions
- Differences across location?
 - Before vs after wastewater treatment
 - Before vs after drinking water treatment
 - Disinfection byproducts



Summary

- NTA is a rapidly growing methodology → standardized QA/QC procedures still under development
- The NTA WebApp is being designed to:
 - Help practitioners more quickly and easily analyze their NTA data
 - Standardize NTA workflows with robust statistical analyses and helpful visualizations
 - Consolidate QC procedures, database querying, and feature identification/prioritization
 - Provide transparency and reproducibility
- De Facto Water Reuse Study data indicate that:
 - Tracers have acceptable mass errors, retention time drift, and CVs
 - 55% of blank-subtracted sample-occurrences present no flags
 - Heat map elucidates spatiotemporal trends and provides guidance for compound ID prioritization

Questions?