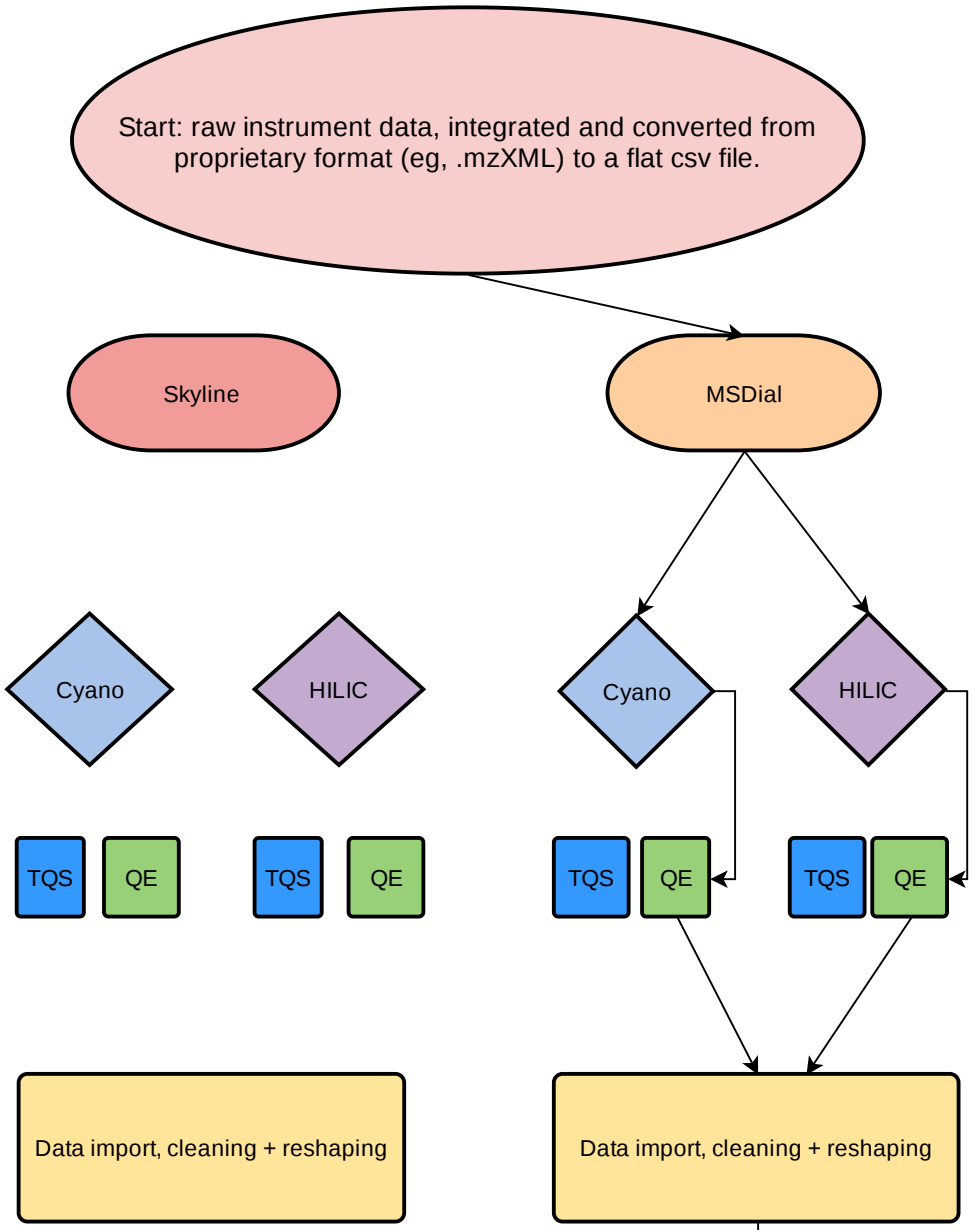
Visual Paradigm Online Diagrams Express Edition



OVERVIEW: MSDial QE

This visual pipeline, and the example pipeline on GitHub, does not cover every step and situation. This is meant to be a guide to the major steps of the targeted processing pipeline.

Determine instrument, software, and run type (QE, TQS, MSDial, Skyline, Cyano, HILIC). This will affect the rest of the analytical process.

The examples shown here are for Cyano and HILIC data that have been run on the QExactive and processed through MSDial.

Throughout the pipeline, files are constantly added to data_processed folder. All raw data is immutable, all created data is disposable.

When modifying, keep track of your changes with version control and clear commits.

DATA IMPORT + CLEANING

Note: The user needs to confirm variables names during the import process to distinguish between HILIC and Cyano runs; ensure that variables are named correctly to access the global environment.

Set header. Remove blank rows, unknown metabolites, and unnecessary columns.
Standardize column names (ie, Metabolite.Name instead of Metabolite.name).
Change column classes to numeric as necessary.
Change dataset orientation from wide to long.
Combine dataset to one dataset in long format instead of four datasets in wide format (Mz, SN, RT, Area).
Remove any "Ingalls_" prefixes that may exist in observations.

Visual Paradigm Online Diagrams Express Edition

Visual Paradigm Online Diagrams Express Edition

Quality Control: Skyline

Quality Control: MSDial

Quality Control + Flagging

Note: The user needs to enter a matching pattern to select the appropriate Cyano or HILIC files from the data_extras folder.

Set parameters for acceptable levels of quality control.
Check run types.
Make retention time table.
Make blank comparison table.
Add signal to noise, area minimum, retention time, and blank comparison flags.
Add comment in csv noting the user-defined parameters.

BMIS

BMIS

Cyano

HILIC

Remove pos + neg HILIC duplicates

BEST-MATCHED INTERNAL STANDARD (B-MIS)

PITFALL: While you will do your best to match the names between the sample key and the data, mistakes happen and it can be very hard to pinpoint where the differences are. Example, one replicate version was QC_1 and one was QC_1a, which was filtered out early on and didn't cause problems until much later...

First: Is BMIS necessary/possible?

Note: User will need to set cut off values for RSD and choose appropriate matching patterns for HILIC and Cyano data.

If working with HILIC data, identify and remove duplicates. Justify your decision.
Identify/visualize internal standards. Look over it to ensure that things are looking more or less "normal".
Using pooled and sample runs, calculate internal standard averages. Create an adjusted area for all pooled and sample runs by comparing each mass feature to each internal standard.
Calculate RSD for each mass feature/internal standard pair, and use this to choose a matched internal standard.
Decide whether to accept a B-MIS or not, according to user-defined values.

Identify + visualize present internal standards, check for normalcy

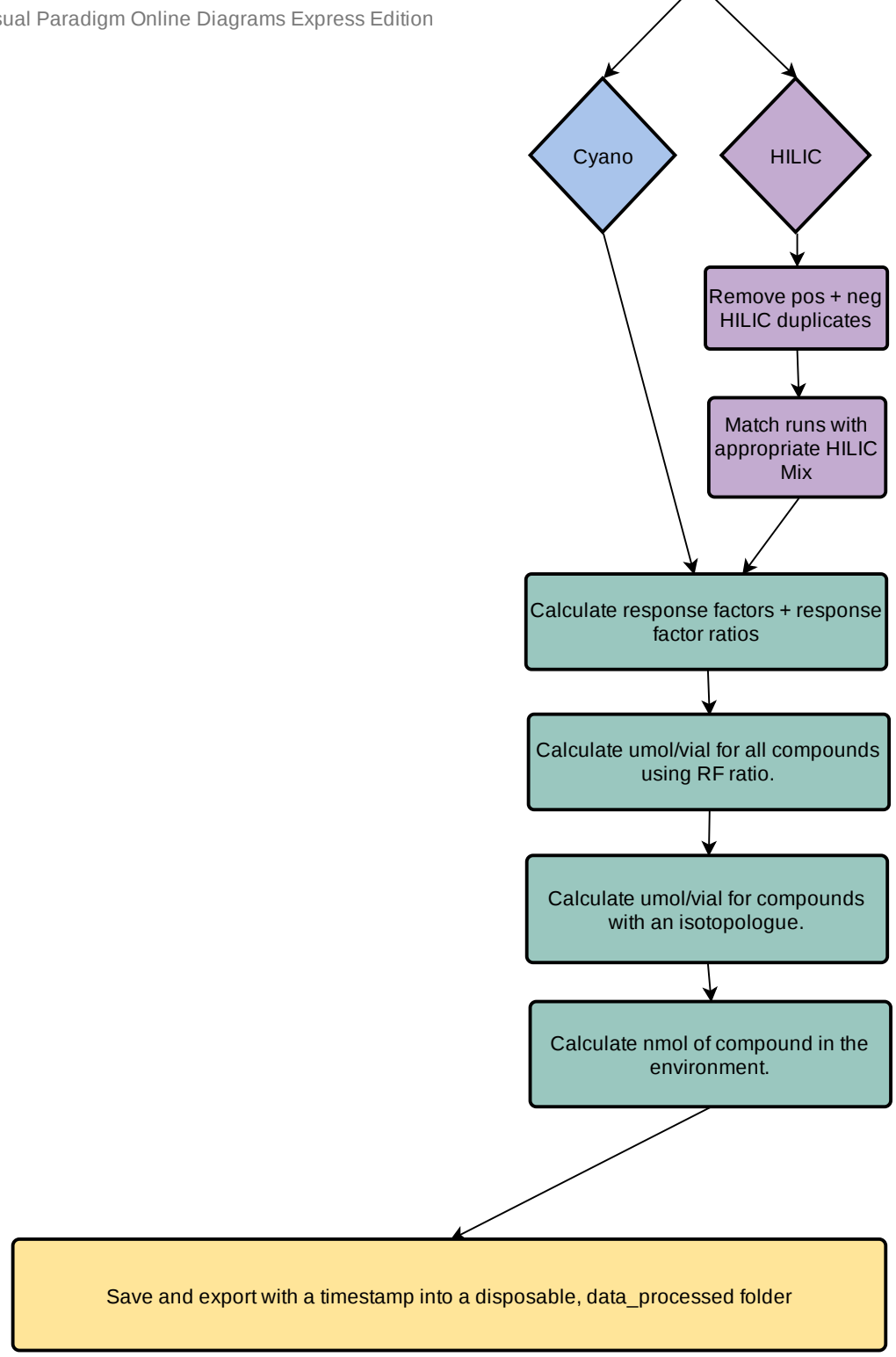
Create adjusted areas, calculate RSD, choose B-MIS or not. Evaluate B-MIS effectiveness

Quantification

Quantification

Visual Paradigm Online Diagrams Express Edition

Visual Paradigm Online Diagrams Express Edition



QUANTIFICATION

PITFALL: In order to quantify using a response factor ratio, you need standards in water, standards in matrix, and water in matrix. The function used to check standard types will not apply to every dataset. There are some similarities, but the user will need to ensure the function is working properly, or adjust it.

PITFALL: Response factor ratios sometimes need to be calculated differently, maybe different runs are being used! Maybe those runs are from different machines!

Again, check if quantification is possible. Quantification requires a response factor ratio or the golden standard of correcting with an isotopologue. Not all runs will have the required data.

Import QC'd files, BMIS'd files, and lab standards.
Again remove duplicates from HILIC QC file.
Combine QC'd data with standards data to obtain mix info, empirical formula, etc.

Check standards run types (see **PITFALL**).
Calculate the response factors and match runs with their appropriate mixes (Mix1, Mix2) HILIC ONLY. Remove internal standards.
Using only standards in water, calculate the RF max and min and the ratio between them. Check for normalcy: the ratio between them should be on the 10 or below order of magnitude.
Using all three run types, calculate the response factor ratios (Standards in matrix - standards in matrix) / (standards in water).
Check for normalcy: The RF ratio should be somewhere on the order of 1.
Supplement your data with information from the Ingalls standards QE.RF ratio IF APPLICABLE.
Calculate the umol/vial for all compounds.
Calculate the umol/vial for compounds that have a matched internal standard (this is the preferable number to have).
Once calculated, replace the umol/vial acquired using response factor with those obtained from using an internal standard calculation.
Calculate the nmol of compound in the environment using the filtered volume, dilution factor, and injection volume.
Obtain numbers of nitrogen and carbon from the compound empirical formulas, and calculate nmol in the environment of each element.
Summarize and present data.

Visual Paradigm Online Diagrams Express Edition