Metabolomics Workbench

mwTab file format specification

Revised:8/30/17

Eoin Fahy, UCSD

First 3 lines of mwTab file

Line1: #METABOLOMICS WORKBENCH
DRCC-curated studies have additional STUDY_ID and ANALYSIS_ID identifiers

Line2: VERSION<13 spaces><TAB>VERSION NUMBER

Line3:CREATED_ON<10 spaces><TAB>DATE in YYYY-MM-DD format

Lines 2 and 3 are 20 characters wide up to the TAB

```
#METABOLOMICS·WORKBENCH·STUDY_ID:ST000001·ANALYSIS_ID:AN000001↵
VERSION············>       1↵
CREATED_ON··········>      2016-09-17↵
#PROJECT↵
PR:PROJECT_TITLE···········>          FatB·Gene·Project↵
PR:PROJECT_TYPE············>          Genotype·treatment↵
```

Project block

#PROJECT
The project block starts with #PROJECT followed by an end-of-line (no spaces, tabs or other characters)

Subsequent lines start with PR: followed by the project metadata item name, for example PR:PROJECT_TITLE The character width before the Tab is 33 characters.

Each metadata item value is placed after the Tab, for example FatB Gene Project
The maximum character width after the Tab is 80 characters. If the metadata item value exceeds 80 characters it is displayed on multiple lines (without using hypenation), each starting with the same metadata item name, for example:
PR:PROJECT_SUMMARY   Experiment to test the consequence of a mutation at the FatB gene (At1g08510)
PR:PROJECT_SUMMARY   the wound-response of Arabidopsis

```
PR:PROJECT_TITLE                 FatB Gene Project
PR:PROJECT_TYPE                  Genotype treatment
PR:PROJECT_SUMMARY               Experiment to test the consequence of a mutation at the FatB gene (At1g08510)
PR:PROJECT_SUMMARY               the wound-response of Arabidopsis
PR:INSTITUTE                     University of California, Davis
PR:DEPARTMENT                    Davis Genome Center
PR:LABORATORY                    Fiehn
PR:LAST_NAME                     Fiehn
PR:FIRST_NAME                    Oliver
PR:ADDRESS                       451 E. Health Sci. Drive, Davis, CA, 95616, USA
PR:EMAIL                         ofiehn@ucdavis.edu
PR:PHONE                         -
PR:PUBLICATIONS                  Quality control for plant metabolomics: reporting MSI-compliant studies.
PR:PUBLICATIONS                  http://www.ncbi.nlm.nih.gov/pubmed/18269577 (PubMed)
```

Study block

#STUDY
The project block starts with #STUDY followed by an end-of-line (no spaces, tabs or other characters)

Subsequent lines start with ST: followed by the study metadata item name, for example ST:STUDY_TITLE The character width before the Tab is 33 characters.

Each metadata item value is placed after the Tab, for example Fatb Induction Experiment (FatBIE)
The maximum character width after the Tab is 80 characters. If the metadata item value exceeds 80 characters it is displayed on multiple lines, each starting with the same metadata item name, for example:
ST:STUDY_SUMMARY   This experiment tests the consequence of a mutation at the FatB gene
ST:STUDY_SUMMARY   in the wound-response of Arabidopsis. The FatB mutant allele (fatb KD J.

```
#STUDY
ST:STUDY_TITLE                  Fatb Induction Experiment (FatBIE)
ST:STUDY_TYPE                   Genotype treatment
ST:STUDY_SUMMARY                This experiment tests the consequence of a mutation at the FatB gene
ST:STUDY_SUMMARY                in the wound-response of Arabidopsis. The FatB mutant allele (fatb KD J.
ST:STUDY_SUMMARY                (Plant Cell 2003, Vol 15, 1020-1033)) was obtained from Dr. Katayonn Dehesh,
ST:STUDY_SUMMARY                of California, Davis, Davis, CA. This allele is in the Ws background.The
ST:STUDY_SUMMARY                growth conditions are as follows: 1. Seeds (between 14 and 16) are sown on
ST:STUDY_SUMMARY                in 100 x 100 x 15mm square Falcon Petri Dishes (Fisher Scientific, catalogue
ST:STUDY_SUMMARY                Seeds were arranged on the plates in a single horizontal line at the 1-cm mark
ST:STUDY_SUMMARY                the top of the plate.2. Each plate contains between 20 and 25-ml of sterile MS
ST:STUDY_SUMMARY                containing 0.1% (w/v) sucrose.3. Prior to sowing, seeds were sterilized by
ST:STUDY_SUMMARY                for 1 minute at room temperature with a 300-l solution of 50% (v/v) ethanol,
ST:STUDY_SUMMARY                solution was removed and replaced with a 300-l solution consisting of 1% (v/v)
ST:STUDY_SUMMARY                20 (Fischer BioReagents, catalogue #BP33750), and 50% (v/v) bleach solution
ST:STUDY_SUMMARY                and incubated at room temperature for 10-minutes. The seeds were then washed
ST:STUDY_SUMMARY                three changes of 0.3-ml of sterile water.
ST:INSTITUTE                    University of California, Davis
ST:DEPARTMENT                   Davis Genome Center
ST:LABORATORY                   Fiehn
ST:LAST_NAME                    Kind
ST:FIRST_NAME                   Tobias
ST:ADDRESS                      451 E. Health Sci. Drive, Davis, CA 95616, USA
ST:EMAIL                        tkind@ucdavis.edu
ST:PHONE                        -
ST:SUBMIT_DATE                  2013-01-15
ST:NUM_GROUPS                   4
ST:TOTAL_SUBJECTS               24
```

Study design (SUBJECT_SAMPLE_FACTORS) block

The study design block starts with

#SUBJECT_SAMPLE_FACTORS: SUBJECT(optional)[tab]SAMPLE[tab]FACTORS(NAME:VALUE pairs separated by |)[tab]Additional sample data

Subsequent lines start with #SUBJECT_SAMPLE_FACTORS <11 spaces> (total of 33 characters in first tab-delimited column)

The 2nd column contains an optional subject identifier for that sample. If no subject identifiers are entered, use a dash (-) in this column.

The 3rd column contains a mandatory sample identifier (e.g. LabF_115873)

The 4th column contains 1 or more NAME/VALUE pairs of experimental factors (conditions) separated by a pipe (|) symbol with a space on either side. The NAME/VALUE items are separated by a colon (:) For example, if the experimental factors are Arabidopsis Genotype and Plant Wounding Treatment, a properly formatted entry would be:

Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded

Sample metadata which varies for every sample or most samples (such as age or BMI) is typically unsuitable for experimental factor designation because it results in too many sample groups, making statistical analysis difficult or impossible. These data should be instead be placed in the 5th column.

The 5th column contains optional additional metadata pertinent to each sample with NAME/VALUE pairs separated by a semicolon and a space (;). The NAME/VALUE items are separated by an 'equal' sign(=) . For example,

Age=47; BMI=28.4

If no additional metadata are entered, leave this column blank.

```
#SUBJECT_SAMPLE_FACTORS:            SUBJECT(optional)[tab]SAMPLE[tab]FACTORS(NAME:VALUE pairs separated by |)[tab]Additional sample data
SUBJECT_SAMPLE_FACTORS              -    LabF_115873    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115878    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115883    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115888    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115893    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115898    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115811    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115816    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115821    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115826    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115831    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115836    Arabidopsis Genotype:Wassilewskija (Ws) | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115904    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115909    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115914    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115919    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115924    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115929    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Control - Non-Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115842    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115847    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115852    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115857    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115862    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
SUBJECT_SAMPLE_FACTORS              -    LabF_115867    Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Wounded
```

SUBJECT, COLLECTION, TREATMENT, SAMPLEPREP, CHROMATOGRAPHY, ANALYSIS ,MS, NMR blocks

The formatting rules for these blocks are the same as for the Project and Study blocks. Each type of block starts with the block name preceded by # and followed by an end-of-line (no spaces, tabs or other characters), e.g:

#SUBJECT

Each subsequent line must start with the appropriate 2-letter identifier followed by a colon:

SUBJECT: SU:

COLLECTION: CO:

TREATMENT: TR:

SAMPLEPREP: SP:

CHROMATOGRAPHY: CH:

ANALYSIS: AN:

MS: MS:

NMR: NM:

In the case of MS analyses, there will be no NMR block. In the case of NMR analyses there will (typically) be no CHROMATOGRAPHY block and no MS block.

```
#MS
MS:INSTRUMENT_NAME . . . . . . . . . . . . .  >    Leco Pegasus III GC-TOF
MS:INSTRUMENT_TYPE . . . . . . . . . . . .  >    GC-TOF
MS:MS_TYPE . . . . . . . . . . . . . . . . .  >    EI
MS:ION_MODE . . . . . . . . . . . . . . . .  >    POSITIVE
```

MS data (results) block

The MS data block starts with #MS_METABOLITE_DATA followed by an end-of-line (no spaces, tabs or other characters)

The 2nd line starts with MS_METABOLITE_DATA:UNITS<Tab> followed by the (mandatory) units type, e.g.:
MS_METABOLITE_DATA:UNITS   peak area

The start of the results section is designated by MS_METABOLITE_DATA_START followed by an end-of-line (no spaces, tabs or other characters) and end of the MS results block is designated by MS_METABOLITE_DATA_END followed by an end-of-line (no spaces, tabs or other characters)

The 1st line after MS_METABOLITE_DATA_START must contain the sample identifiers which should be identical to those given in the Study design section. The 1st column of this line contains the word "Samples" and subsequent columns contain the sample identifiers.

The 2nd line optionally contains the name/value pairs of the corresponding experimental factors specified in the Study design section, with the word "Factors" in the 1st column. This line may be omitted.

Subsequent lines contain the results with the unique metabolite (feature) name in the 1st column followed by measurements for each listed sample.

```
#MS_METABOLITE_DATA
MS_METABOLITE_DATA:UNITS              peak area
MS_METABOLITE_DATA_START
Samples LabF_115904      LabF_115909      LabF_115914      LabF_115919      LabF_11592
Factors Arabidopsis Genotype:fatb-ko KD; At1g08510 | Plant Wounding Treatment:Cont
1_2_4-benzenetriol      1874.0000      3566.0000      1945.0000      1456.0000
1-monostearin   987.0000      450.0000      1910.0000      549.0000      10
2-hydroxyvaleric acid   771.0000      931.0000      1114.0000      509.0000
...
MS_METABOLITE_DATA_END
```

MS metabolite metadata block

The MS metabolite metadata block starts with #METABOLITES followed by an end-of-line (no spaces, tabs or other characters)

The start of the metabolite metadata section is designated by METABOLITES_START followed by an end-of-line (no spaces, tabs or other characters) and end is designated by METABOLITES_END followed by an end-of-line (no spaces, tabs or other characters)

The 1st line after METABOLITES_START must contain metabolite metadata headings. The first column of this line contains the word "metabolite_name". The metabolite metadata headings may be specified by submitter. Typically these may include m/z, quantitated m/z, retention time, retention index, PubChem compound ID, InChiKey, etc.
Subsequent lines contain the metabolite metadata with the unique metabolite (feature) name in the 1st column followed by values for each listed heading. Metabolite (feature) names must exactly match those listed in the previous MS_METABOLITE_DATA block.
If no metabolite metadata is being supplied, this section must contain the metabolite (feature) names in the 1st column (and no other columns)

```
#METABOLITES
METABOLITES_START
metabolite_name moverz_quant      retention_index pubchem_id
1,2,4-benzenetriol        239      522741  10787
1-monostearin   399       959625  107036
2-hydroxyvaleric acid     131      310750  98009
3-phosphoglycerate        299      611619  724
...
METABOLITES_END
```

NMR binned data block

The NMR binned data block starts with #NMR_BINNED_DATA followed by an end-of-line (no spaces, tabs or other characters)

The start of the results section is designated by NMR_BINNED_DATA_START followed by an end-of-line (no spaces, tabs or other characters) and end of the results block is designated by NMR_BINNED_DATA_END followed by an end-of-line (no spaces, tabs or other characters)

The 1st line after NMR_BINNED_DATA_START must contain the sample identifiers which should be identical to those given in the Study design section. The 1st column of this line contains the word "Bin range(ppm)" and subsequent columns contain the sample identifiers.
Subsequent lines contain the results with the bin range in the 1st column followed by measurements for each listed sample.
The 2 values in each bin range are separated by 3 periods (...)

```
#NMR_BINNED_DATA
NMR_BINNED_DATA_START
Bin range(ppm)   s01      s02      s03      s04      s05
0.50...0.54      -0.2341 -0.1662 -0.1521 -0.2534 -0.2836
0.54...0.58      -0.2993 -0.2433 -0.2548 -0.2315 -0.1850
0.58...0.62      -0.1697 -0.0500 -0.0600 -0.2042 -0.3757
...
NMR_BINNED_DATA_END
```

Project fields

PR:PROJECT_TITLE
PR:PROJECT_TYPE
PR:PROJECT_SUMMARY
PR:INSTITUTE
PR:DEPARTMENT
PR:LABORATORY
PR:LAST_NAME
PR:FIRST_NAME
PR:ADDRESS
PR:EMAIL
PR:PHONE
PR:FUNDING_SOURCE
PR:PROJECT_COMMENTS
PR:PUBLICATIONS
PR:CONTRIBUTORS
PR:DOI (added by DRCC only)

(required fields in red)

Study fields

ST:STUDY_TITLE
ST:STUDY_TYPE
ST:STUDY_SUMMARY
ST:INSTITUTE
ST:DEPARTMENT
ST:LABORATORY
ST:LAST_NAME
ST:FIRST_NAME
ST:ADDRESS
ST:EMAIL
ST:PHONE
ST:NUM_GROUPS
ST:TOTAL_SUBJECTS
ST:NUM_MALES
ST:NUM_FEMALES
ST:STUDY_COMMENTS
ST:PUBLICATIONS

(required fields in red)

Subject fields

SU:SUBJECT_TYPE
SU:SUBJECT_SPECIES
SU:TAXONOMY_ID
SU:GENOTYPE_STRAIN
SU:AGE_OR_AGE_RANGE
SU:WEIGHT_OR_WEIGHT_RANGE
SU:HEIGHT_OR_HEIGHT_RANGE
SU:GENDER
SU:HUMAN_RACE
SU:HUMAN_ETHNICITY
SU:HUMAN_TRIAL_TYPE
SU:HUMAN_LIFESTYLE_FACTORS
SU:HUMAN_MEDICATIONS
SU:HUMAN_PRESCRIPTION_OTC
SU:HUMAN_SMOKING_STATUS
SU:HUMAN_ALCOHOL_DRUG_USE

SU:HUMAN_NUTRITION
SU:HUMAN_INCLUSION_CRITERIA
SU:HUMAN_EXCLUSION_CRITERIA
SU:ANIMAL_ANIMAL_SUPPLIER
SU:ANIMAL_HOUSING
SU:ANIMAL_LIGHT_CYCLE
SU:ANIMAL_FEED
SU:ANIMAL_WATER
SU:ANIMAL_INCLUSION_CRITERIA
SU:CELL_BIOSOURCE_OR_SUPPLIER
SU:CELL_STRAIN_DETAILS
SU:SUBJECT_COMMENTS
SU:CELL_PRIMARY_IMMORTALIZED
SU:CELL_PASSAGE_NUMBER
SU:CELL_COUNTS
SU:SPECIES_GROUP

(required fields in red)

Collection fields

CO:COLLECTION_SUMMARY
CO:COLLECTION_PROTOCOL_ID
CO:COLLECTION_PROTOCOL_FILENAME
CO:COLLECTION_PROTOCOL_COMMENTS
CO:SAMPLE_TYPE
CO:COLLECTION_METHOD
CO:COLLECTION_LOCATION
CO:COLLECTION_FREQUENCY
CO:COLLECTION_DURATION
CO:COLLECTION_TIME
CO:VOLUMEORAMOUNT_COLLECTED
CO:STORAGE_CONDITIONS
CO:COLLECTION_VIALS
CO:STORAGE_VIALS
CO:COLLECTION_TUBE_TEMP
CO:ADDITIVES
CO:BLOOD_SERUM_OR_PLASMA
CO:TISSUE_CELL_IDENTIFICATION
CO:TISSUE_CELL_QUANTITY_TAKEN

(required fields in red)

Treatment fields

TR:TREATMENT_SUMMARY
TR:TREATMENT_PROTOCOL_ID
TR:TREATMENT_PROTOCOL_FILENAME
TR:TREATMENT_PROTOCOL_COMMENTS
TR:TREATMENT
TR:TREATMENT_COMPOUND
TR:TREATMENT_ROUTE
TR:TREATMENT_DOSE
TR:TREATMENT_DOSEVOLUME
TR:TREATMENT_DOSEDURATION
TR:TREATMENT_VEHICLE
TR:ANIMAL_VET_TREATMENTS
TR:ANIMAL_ANESTHESIA
TR:ANIMAL_ACCLIMATION_DURATION
TR:ANIMAL_FASTING
TR:ANIMAL_ENDP_EUTHANASIA
TR:ANIMAL_ENDP_TISSUE_COLL_LIST
TR:ANIMAL_ENDP_TISSUE_PROC_METHOD
TR:ANIMAL_ENDP_CLINICAL_SIGNS

TR:HUMAN_FASTING
TR:HUMAN_ENDP_CLINICAL_SIGNS
TR:CELL_STORAGE
TR:CELL_GROWTH_CONTAINER
TR:CELL_GROWTH_CONFIG
TR:CELL_GROWTH_RATE
TR:CELL_INOC_PROC
TR:CELL_MEDIA
TR:CELL_ENVIR_COND
TR:CELL_HARVESTING
TR:PLANT_GROWTH_SUPPORT
TR:PLANT_GROWTH_LOCATION
TR:PLANT_PLOT_DESIGN
TR:PLANT_LIGHT_PERIOD
TR:PLANT_HUMIDITY
TR:PLANT_TEMP
TR:PLANT_WATERING_REGIME
TR:PLANT_NUTRITIONAL_REGIME
TR:PLANT_ESTAB_DATE

TR:PLANT_HARVEST_DATE
TR:PLANT_GROWTH_STAGE
TR:PLANT_METAB_QUENCH_METHOD
TR:PLANT_HARVEST_METHOD
TR:PLANT_STORAGE
TR:CELL_PCT_CONFLUENCE
TR:CELL_MEDIA_LASTCHANGED

(required fields in red)

Sampleprep fields

SP:SAMPLEPREP_SUMMARY
SP:SAMPLEPREP_PROTOCOL_ID
SP:SAMPLEPREP_PROTOCOL_FILENAME
SP:SAMPLEPREP_PROTOCOL_COMMENTS
SP:PROCESSING_METHOD
SP:PROCESSING_STORAGE_CONDITIONS
SP:EXTRACTION_METHOD
SP:EXTRACT_CONCENTRATION_DILUTION
SP:EXTRACT_ENRICHMENT
SP:EXTRACT_CLEANUP
SP:EXTRACT_STORAGE
SP:SAMPLE_RESUSPENSION
SP:SAMPLE_DERIVATIZATION
SP:SAMPLE_SPIKING
SP:ORGAN
SP:ORGAN_SPECIFICATION
SP:CELL_TYPE
SP:SUBCELLULAR_LOCATION

(required fields in red)

Chromatography fields

CH:CHROMATOGRAPHY_SUMMARY
CH:CHROMATOGRAPHY_TYPE
CH:INSTRUMENT_NAME
CH:COLUMN_NAME
CH:FLOW_GRADIENT
CH:FLOW_RATE
CH:COLUMN_TEMPERATURE
CH:METHODS_FILENAME
CH:SOLVENT_A
CH:SOLVENT_B
CH:METHODS_ID
CH:COLUMN_PRESSURE
CH:INJECTION_TEMPERATURE
CH:INTERNAL_STANDARD
CH:INTERNAL_STANDARD_MT
CH:RETENTION_INDEX
CH:RETENTION_TIME
CH:SAMPLE_INJECTION
CH:SAMPLING_CONE
CH:ANALYTICAL_TIME

CH:CAPILLARY_VOLTAGE
CH:MIGRATION_TIME
CH:OVEN_TEMPERATURE
CH:PRECONDITIONING
CH:RUNNING_BUFFER
CH:RUNNING_VOLTAGE
CH:SHEATH_LIQUID
CH:TIME_PROGRAM
CH:TRANSFERLINE_TEMPERATURE
CH:WASHING_BUFFER
CH:WEAK_WASH_SOLVENT_NAME
CH:WEAK_WASH_VOLUME
CH:STRONG_WASH_SOLVENT_NAME
CH:STRONG_WASH_VOLUME
CH:TARGET_SAMPLE_TEMPERATURE
CH:SAMPLE_LOOP_SIZE
CH:SAMPLE_SYRINGE_SIZE
CH:RANDOMIZATION_ORDER
CH:CHROMATOGRAPHY_COMMENTS

(required fields in red)

Analysis fields

AN:ANALYSIS_TYPE
AN:LABORATORY_NAME
AN:OPERATOR_NAME
AN:DETECTOR_TYPE
AN:SOFTWARE_VERSION
AN:ACQUISITION_DATE
AN:ANALYSIS_PROTOCOL_FILE
AN:ACQUISITION_PARAMETERS_FILE
AN:PROCESSING_PARAMETERS_FILE
AN:DATA_FORMAT

(required fields in red)

MS fields

MS:INSTRUMENT_NAME
MS:INSTRUMENT_TYPE
MS:MS_TYPE
MS:ION_MODE
MS:MS_COMMENTS
MS:CAPILLARY_TEMPERATURE
MS:CAPILLARY_VOLTAGE
MS:COLLISION_ENERGY
MS:COLLISION_GAS
MS:DRY_GAS_FLOW
MS:DRY_GAS_TEMP
MS:FRAGMENT_VOLTAGE
MS:FRAGMENTATION_METHOD
MS:GAS_PRESSURE
MS:HELIUM_FLOW
MS:ION_SOURCE_TEMPERATURE
MS:ION_SPRAY_VOLTAGE
MS:IONIZATION
MS:IONIZATION_ENERGY
MS:IONIZATION_POTENTIAL

MS:MASS_ACCURACY
MS:PRECURSOR_TYPE
MS:REAGENT_GAS
MS:SOURCE_TEMPERATURE
MS:SPRAY_VOLTAGE
MS:ACTIVATION_PARAMETER
MS:ACTIVATION_TIME
MS:ATOM_GUN_CURRENT
MS:AUTOMATIC_GAIN_CONTROL
MS:BOMBARDMENT
MS:CDL_SIDE_OCTOPOLES_BIAS_VOLTAGE
MS:CDL_TEMPERATURE
MS:DATAFORMAT
MS:DESOLVATION_GAS_FLOW
MS:DESOLVATION_TEMPERATURE
MS:INTERFACE_VOLTAGE
MS:IT_SIDE_OCTOPOLES_BIAS_VOLTAGE
MS:LASER
MS:MATRIX

MS:NEBULIZER
MS:OCTPOLE_VOLTAGE
MS:PROBE_TIP
MS:RESOLUTION_SETTING
MS:SAMPLE_DRIPPING
MS:SCAN_RANGE_MOVERZ
MS:SCANNING
MS:SCANNING_CYCLE
MS:SCANNING_RANGE
MS:SKIMMER_VOLTAGE
MS:TUBE_LENS_VOLTAGE
MS:MS_RESULTS_FILE

(required fields in red)

NMR fields

NM:INSTRUMENT_NAME
NM:INSTRUMENT_TYPE
NM:NMR_EXPERIMENT_TYPE
NM:NMR_COMMENTS
NM:FIELD_FREQUENCY_LOCK
NM:STANDARD_CONCENTRATION
NM:SPECTROMETER_FREQUENCY
NM:NMR_PROBE
NM:NMR_SOLVENT
NM:NMR_TUBE_SIZE
NM:SHIMMING_METHOD
NM:PULSE_SEQUENCE
NM:WATER_SUPPRESSION
NM:PULSE_WIDTH
NM:POWER_LEVEL
NM:RECEIVER_GAIN
NM:OFFSET_FREQUENCY
NM:PRESATURATION_POWER_LEVEL
NM:CHEMICAL_SHIFT_REF_CPD

NM:TEMPERATURE
NM:NUMBER_OF_SCANS
NM:DUMMY_SCANS
NM:ACQUISITION_TIME
NM:RELAXATION_DELAY
NM:SPECTRAL_WIDTH
NM:NUM_DATA_POINTS_ACQUIRED
NM:REAL_DATA_POINTS
NM:LINE_BROADENING
NM:ZERO_FILLING
NM:APODIZATION
NM:BASELINE_CORRECTION_METHOD
NM:CHEMICAL_SHIFT_REF_STD
NM:BINNED_INCREMENT
NM:BINNED_DATA_NORMALIZATION_METHOD   -
NM:BINNED_DATA_PROTOCOL_FILE
NM:BINNED_DATA_CHEMICAL_SHIFT_RANGE   -
NM:BINNED_DATA_EXCLUDED_RANGE

(required fields in red)

Order of metadata and data blocks (MS)

Results from targeted experiments with named metabolites are entered directly in the mwTab file within the MS_METABOLITE_DATA and METABOLITES blocks

```
#METABOLOMICS WORKBENCH STUDY_ID:ST000001 ANALYSIS_ID:AN000001
VERSION                        1
CREATED_ON                     2016-09-17
#PROJECT
...
#STUDY
...
#SUBJECT
...
#SUBJECT_SAMPLE_FACTORS:                     SUBJECT(optional)[tab]SAMP
...
#COLLECTION
...
#TREATMENT
...
#SAMPLEPREP
...
#CHROMATOGRAPHY
...
#ANALYSIS
...
#MS
...
#MS_METABOLITE_DATA
MS_METABOLITE_DATA:UNITS           peak area
MS_METABOLITE_DATA_START
...
MS_METABOLITE_DATA_END
#METABOLITES
METABOLITES_START
...
METABOLITES_END
#END
```

Order of metadata and data blocks (MS)

Results from untargeted experiments (usually 1000's of features) are referenced by the "MS_RESULTS_FILE" field in the MS block of the mwTab file. The results file should be a tab-delimited text file containing sample names identical to those listed in the Study design (SUBJECT_SAMPLE_FACTORS) section

```
#METABOLOMICS WORKBENCH
VERSION                          1
CREATED_ON                      August 30, 2017, 9:09 am
#PROJECT
...
#STUDY
...
#SUBJECT
...
#SUBJECT_SAMPLE_FACTORS:                    SUBJECT(optional)[tab]SAMPLE[tab
...
#COLLECTION
...
#TREATMENT
...
#SAMPLEPREP
...
#CHROMATOGRAPHY
...
#ANALYSIS
...
#MS
...
#MS_RESULTS_FILE                    Study_A17B.txt  UNITS:Peak height
#END
```

Specify units of measurement

Format: MS:MS_RESULTS_FILE<15 spaces><TAB><file name><TAB>UNITS:<units of measurement>

Order of metadata and data blocks (NMR)

Note the absence of a chromatography block

```
#METABOLOMICS WORKBENCH
VERSION                          1
CREATED_ON                       2016-09-17
#PROJECT
...
#STUDY
...
#SUBJECT
...
#SUBJECT_SAMPLE_FACTORS:                    SUBJECT(optional)
...
#COLLECTION
...
#TREATMENT
...
#SAMPLEPREP
...
#ANALYSIS
...
#NMR
...
#NMR_METABOLITE_DATA
NMR_METABOLITE_DATA:UNITS          uM
NMR_METABOLITE_DATA_START
...
NMR_METABOLITE_DATA_END
#METABOLITES
METABOLITES_START
...
METABOLITES_END
#END
```