mzTab cheat sheet

Data types

Type	Format

Parameter	[{CV label}, {accession}, {name}, {value}]
	`[MS, MS:1001171, Mascot:score, 40.21]`
	`[,,A user parameter, The value]`

Parameter List	'|' separated `[MS, MS:1001171, Mascot:score, 40.21]|[,,Another parameter,]`
String List	'|' separated `first string|second string`

Modification	{position}{reliability score}-{Modification or Substitution identifier}{Neutral loss}
	`1[0.5]-UNIMOD:4`
	`1-UNIMOD:4|[MS, MS:1001524, fragment neutral loss, 63.998285]`
	`3|4-MOD:00412,8-MOD:00412`
	`10[MS,MS:1001876, modification probability, 0.8]-CHEMMOD:+57.0215`
	`23-CHEMMOD:+NH4`
	`5-SUBST:R`

Spec Ref	ms_file[1-n]:{SPEC REF} `ms_file[2]:index=7|ms_file[2]:index=9`

MTD - Metadata section (Key-value)

Field	Type	Description
{UNIT_ID}-title	String	The unit's title
{UNIT_ID}-description	String	The unit's description
{UNIT_ID}-sample_processing[1-n]	Parameter List	Description of the sample processing.
{UNIT_ID}-instrument[1-n]-source	Parameter	The instrument's source
{UNIT_ID}-instrument[1-n]-analyzer	Parameter	The instrument's analyzer
{UNIT_ID}-instrument[1-n]-detector	Parameter	The instrument's detector
{UNIT_ID}-software[1-n]	Parameter	Analysis software used in the order it was used.
{UNIT_ID}-software[1-n]-setting	String	Software setting used.
{UNIT_ID}-false_discovery_rate	Parameter List	False discovery rate(s)for the experiment.
{UNIT_ID}-publication	String List	Publication ids (pubmed / doi).
{UNIT_ID}-contact[1-n]-name	String	Contact name.
{UNIT_ID}-contact[1-n]-affiliation	String	Contact affiliation.
{UNIT_ID}-contact[1-n]-email	String	Contact's e-mail address.
{UNIT_ID}-uri	URI	Points to the unit's source data.
{UNIT_ID}-mod	Parameter List	Modifications reported in the unit.
{UNIT_ID}-mod-probability_method	Parameter	Method used to report modification probabilities.
{UNIT_ID}-quantification_method	Parameter	Quantification method used.
{UNIT_ID}-protein-quantification_unit	Parameter	Unit of protein quantification results.
{UNIT_ID}-peptide-quantification_unit	Parameter	Unit of peptide quantification results.
{UNIT_ID}-ms_file[1-n]-format	Parameter	Data format of the external MS data file.
{UNIT_ID}-ms_file[1-n]-location	URI	Location of the external MS data file.
{UNIT_ID}-ms_file[1-n]-id_format	Parameter	Identifier format of the external MS data file.
{UNIT_ID}-custom	Parameter	Additional parameters.
{UNIT_ID}(-{SUB_ID})-species[1-n]	Parameter	Species of the unit / subsample.
{UNIT_ID}(-{SUB_ID})-tissue[1-n]	Parameter	Tissue of the unit / subsample.
{UNIT_ID}(-{SUB_ID})-cell_type[1-n]	Parameter	Cell type of the unit / subsample.
{UNIT_ID}(-{SUB_ID})-disease[1-n]	Parameter	Disease state of the unit / subsample.
{UNIT_ID}-{SUB_ID}-description	String	Description of the subsample.
{UNIT_ID}-{SUB_ID}-quantification_reagent	Parameter	Quantification reagent used to label the subsample.
{UNIT_ID}-{SUB_ID}-custom	Parameter	Additional parameters for the subsample.

PRH/PRT - Protein section (Table based)		
Column	**Type**	**Description**
accession	String	The protein's accession.
unit_id	String	The unit's id.
description	String	Human readable description (i.e. the name)
taxid	Integer	NEWT taxonomy for the species.
species	String	Human readable species
database	String	Name of the protein database.
database_version	String	Version of the protein database.
search_engine	Parameter List	Search engine(s) identifying the protein.
search_engine_score	Parameter List	Search engine(s) reliability score(s).
reliability	Integer (1-3)	Identification reliability.
num_peptides	Integer	Number of PSMs assigned to the protein.
num_peptides_distinct	Integer	Distinct (sequence + modifications) # of peptides.
num_peptides_unambiguous	Integer	Distinct number of unambiguous peptides.
ambiguity_members	String List (",")	Alternative protein identifications.
modifications	Modification List (",")	Modifications identified in the protein.
uri	URI	Location of the protein's source entry.
go_terms	String List	List of GO terms for the protein.
protein_coverage	Double (0-1)	Amount of protein sequence identified.
protein_abundance_sub[1-n]	Double	Protein abundance in the subsample.
protein_abundance_stdev_sub[1-n]	Double	Standard deviation of the protein abundance.
protein_abundance_std_error_sub[1-n]	Double	Standard error of the protein abundance.
opt_*	Column	Additional columns must start with "opt_"

PEH/PEP - Peptide section (Table based)		
Column	**Type**	**Description**
sequence	String	The peptide's sequence.
accession	String	The protein's accession.
unit_id	String	The unit's id.
unique	Boolean (1/0)	Peptide is unique for the protein.
database	String	Name of the sequence database.
database_version	String	Version (and optionally # of entries).
search_engine	Parameter List	Search engine(s) that identified the peptide.
search_engine_score	Parameter List	Search engine(s) score(s) for the peptide.
reliability	Integer(1-3)	Identification reliability for the peptide.
modifications	Modification List	Modifications identified in the peptide.
retention_time	Double List ("|")	Time points in seconds. Semantics may vary.
charge	Integer	Precursor ion's charge.
mass_to_charge	Double	Precursor ion's *m/z*.
uri	URI	Location of the PSM's source entry.
spectra_ref	Spec Ref	Spectra identifying the peptide.
peptide_abundance_sub[1-n]	Double	Peptide abundance in the subsample.
peptide_abundance_stdev_sub[1-n]	Double	Peptide abundance standard deviation.
peptide_abundance_std_error_sub[1-n]	Double	Peptide abundance standard error.
opt_*	Column	Optional columns must start with "opt_".

SMH/SML Small molecule section (Table based)		
Column	**Type**	**Description**
identifier	String	The small molecule's identifier.
unit_id	String	The unit's id.
chemical_formula	String	Chemical formula of the identified compound.
smiles	String	Molecular structure in SMILES format.
inchi_key	String	InChi Key of the identified compound.
description	String	Human readable description (i.e. the name)
mass_to_charge	Double	The precursor ion's *m/z*.
charge	Integer	The precursor ion's charge.
retention_time	Double List ("|")	List of time points in seconds.
taxid	Integer	NEWT taxonomy id of the species.
species	String	Human readable name of the species.
database	String	Name of the used database.
database_version	String	Version (and optionally # of compounds)
reliability	Integer (1-3)	The identification reliability.
uri	URI	The source entry's location.
spectra_ref	Spec Ref	References to spectra identify the small molecule.
search_engine	Parameter List	Search engine(s) identifying the small molecule.
search_engine_score	Parameter List	Search engine(s) identifications score(s).
modifications	Modification List	Modifications identified on the small molecule.
smallmolecule_abundance_sub[1-n]	Double	Abundance in the subsample.
smallmolecule _abundance_stdev_sub[1-n]	Double	Standard deviation of the abundance.
smallmolecule _abundance_std_error_sub[1-n]	Double	Standard error of the abundance.
opt_*	Column	Optional columns must start with "opt_"