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 identifier}  1(0.5)-UNIMOD:4  10(0.8)|11(0.2)-CHEMMOD:+57.0215 23-CHEMMOD:+NA4 |
| 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}-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 | Double | 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 | Double | | 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\_” |
|  |  | |  |