**mzQuantML documentation**

Basic file structure is as follows, more detail on each below:

* mzQuantML – root element
  + ProteinList – quantified proteins and/or protein groups
  + PeptideList – quantified peptides
  + AssayList –lists of “assays” capturing raw files, labelling etc
  + FeatureList – features on individual spectra
  + QuantificationList – quantified values extracted from the features (inc ratios of two features)
  + DataProcessingList – protocols used

**ProteinList**

* QuantProteinAmbiguityGroup
  + Identification reference for the group (e.g. a PAG within mzIdentML)
  + GroupedQuantitiationResult – Measurement or ratio of this protein group within one study variable e.g. across a range of technical replicates
  + QuantitationResult – measure or ratio of this protein group within one condition. Reference to assay(s) - (usually one raw file, one or more assays, where different assays capture the different labels if applicable)
  + GlobalQuantitationResult – stats about this result e.g. p-values?
  + QuantProteinDetectionHypothesis
* QuantProteinDetectionHypothesis
  + Identification reference for the protein
  + GroupedQuantitiationResult – Measurement or ratio of this protein within one study variable e.g. across a range of technical replicates
  + QuantitationResult – measure or ratio of this protein within one condition (assay)
  + PeptideRefs – references to peptides below

**PeptideList**

* Peptide
  + Sequence
  + IdentificationReference (e.g. SIIs in mzIdentML)
  + Modifications
  + GroupedQuantitationResult – measurement or ratio in one study variable
  + QuantitationResult – result for one assay

**AssayList**

* Assay – one experimental “run” e.g. label-free replicate, one channel of iTRAQ, one SILAC label
  + RawFile(s) – (presumably) multiple raw files allowed here in case the “run” actually combines raw files, e.g. if there was pre-separation
  + Label (0..1) – cvParam of label/channel
  + IdentificationFiles – e.g. ref to mzIdentML file(s) associated with this assay
  + DataProcessingOrder – *TO DO documentation, this has been discussed on list*
  + cvParam

**FeatureList**

* Feature – one feature within (typically an MS1) spectrum/scan
  + @Attributes – RT, MZ, intensity and charge (intensity stores the raw intensity e.g. to be used in a quantificationUnit)
  + massTrace (encoding the coordinates of the feature)
    - RTStart
    - RTStop
    - MZStart
    - MZStop
  + IdentificationReferences

**QuantificationList**

* QuantificationUnit
  + @Attributes – id, name, value; value stores the value of the quantified feature(s)
  + FeatureReferences
    - GroupedFeatureRefs – captures references to sets of features over which a mathematical operation has been performed e.g. averaging over replicates
* Ratio
  + References to two quantification units used as denominator and numerator
* Measurement – purpose of measurement element was to capture a similar concept to quantitation unit – this can probably be removed

**DataProcessingList**

* dataProcessing
  + processingMethod - captures cvParams for this method of data processing, to be referenced elsewhere in the file
    - cvParam

**Outstanding questions:**

1. StudyVariables are in the schema but don’t seem to be reachable from the root

2. Linkage between grouped values in level 1 and level 2 of the schema, seem to be multiple places for summing/averaging over replicates