Mathias Walzer, University of Tübingen, Germany
Oliver Kohlbacher, University of Tübingen, Germany
Florian Reisinger, European Bioinformatics Institute
J. Alberto Medina-Aunon, Centro Nacional de Biotecnología, Madrid, Spain
Julian Uszkoreit, Medizinisches Proteom-Center, Ruhr-Universität Bochum
Gerhard Mayer, Medizinisches Proteom-Center, Ruhr-Universität Bochum
Martin Eisenacher, Medizinisches Proteom-Center, Ruhr-Universität Bochum
Andrew R. Jones, University of Liverpool

Feb 2013

mzQuantML: exchange format for quantitation values associated with peptides, proteins and small molecules from mass spectra

Status of This Document

This document presents a final specification for the mzQuantML data format developed by the HUPO Proteomics Standards Initiative. Distribution is unlimited.

Version of This Document
The current version of this document is: version 1.0.0

Abstract

The Human Proteome Organisation (HUPO) Proteomics Standards Initiative (PSI) defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification. The Proteomics Informatics Working Group is developing standards for describing the results of identification and quantitation processes for proteins, peptides and protein modifications from mass spectrometry. This document defines an XML schema that can be used to describe the outputs of quantitation software for proteomics. Limited support is also provided for capturing quantitation values about small molecule generated from mass spectrometry.

Contents

Abstract....1
1. Introduction....3
1.1 Background....3
1.2 Document Structure....4
2. Use Cases for mzQuantML....4
3. Concepts and Terminology....5
4. Relationship to Other Specifications....5
4.1 Important concepts from FuGE....6
4.2 The PSI Mass Spectrometry Controlled Vocabulary (CV)....6
4.3 Validation of controlled vocabulary terms....7
4.3.1 Validation of values used within CvParams....7
5. Resolved Design and scope issues....8
5.1.1 Handling updates to the controlled vocabulary....8
5.1.2 Use of mzQuantML for analysis pipelines....8
5.2 Encoding zeroes, nulls, infinity and calculation errors....8
5.3 Protein grouping....9
5.4 Comments on Specific Use Cases....9
5.4.1 MS1 label-free intensity....10
5.4.2 MS1 label-based....10

5.4.3 MS2 spectral counting 11
5.4.4 MS2 tag-based 11
5.5 Semantic validation rules 12
5.6 Other supporting materials 12
5.7 Open Issues 12
5.7.1 Support for metabolomics 12
6. Model in XML Schema 12
6.1 Element <MzQuantML> 13
6.2 Element <Affiliation> 16
6.3 Element <AnalysisSummary> 16
6.4 Element <Assay> 17
6.5 Element <AssayList> 18
6.6 Element <AssayQuantLayer> 19
6.7 Element <Assay_refs> 20
6.8 Element <AuditCollection> 20
6.9 Element <BibliographicReference> 21
6.10 Element <Column> 22
6.11 Element <ColumnDefinition> 22
6.12 Element <ColumnIndex> 23
6.13 Element <ContactRole> 23
6.14 Element <Cv> 23
6.15 Element <CvList> 24
6.16 Element <cvParam> 24
6.17 Element <DatabaseName> 25
6.18 Element <DataMatrix> 25
6.19 Element <DataProcessing> 26
6.20 Element <DataProcessingList> 26
6.21 Element <DataType> 27
6.22 Element <DBIdentificationRef> 33
6.23 Element <DenominatorDataType> 33
6.24 Element <EvidenceRef> 34
6.25 Element <ExternalFormatDocumentation> 35
6.26 Element <Feature> 35
6.27 Element <FeatureList> 36
6.28 Element <FeatureQuantLayer> 39
6.29 Element <Feature_refs> 39
6.30 Element <FileFormat> 39
6.31 Element <GlobalQuantLayer> 40
6.32 Element <IdentificationFile> 41
6.33 Element <IdentificationFiles> 42
6.34 Element <IdentificationFile_refs> 42
6.35 Element <IdentificationRef> 42
6.36 Element <InputFiles> 43
6.37 Element <InputObject_refs> 45
6.38 Element <Label> 45
6.39 Element <MassTrace> 45
6.40 Element <MethodFile> 46
6.41 Element <MethodFiles> 46
6.42 Element <Modification> 47
6.42.1 Specification of a <Modification> searched for 47
6.42.2 Modifications used to define a label or tag for quantification 47
6.42.2.1 For multiplex techniques, such as a 4plex iTRAQ analysis 47
6.42.2.2 For a 2plex SILAC, experiment 48
6.42.2.3 In label-free analyses 48
6.42.3 Modifications in Small Molecules 48
6.43 Element <MS2AssayQuantLayer> 48
6.44 Element <MS2RatioQuantLayer> 49

6.45 Element <MS2StudyVariableQuantLayer> ... 49
6.46 Element <NumeratorDataType> ... 50
6.47 Element <Organization> ... 50
6.48 Element <OutputObject_refs> ... 51
6.49 Element <ParentOrganization> ... 51
6.50 Element <PeptideConsensus> ... 51
6.51 Element <PeptideConsensusList> ... 53
6.52 Element <PeptideConsensus_refs> ... 56
6.53 Element <PeptideSequence> ... 56
6.54 Element <Person> ... 57
6.55 Element <ProcessingMethod> ... 57
6.56 Element <Protein> ... 58
6.57 Element <ProteinGroup> ... 59
6.58 Element <ProteinGroupList> ... 60
6.59 Element <ProteinList> ... 63
6.60 Element <ProteinRef> ... 65
6.61 Element <Provider> ... 65
6.62 Element <Ratio> ... 66
6.63 Element <RatioCalculation> ... 67
6.64 Element <RatioList> ... 67
6.65 Element <RatioQuantLayer> ... 68
6.66 Element <RawFile> ... 69
6.67 Element <RawFilesGroup> ... 70
6.68 Element <Role> ... 71
6.69 Element <Row> ... 71
6.70 Element <SearchDatabase> ... 71
6.71 Element <SmallMolecule> ... 72
6.72 Element <SmallMoleculeList> ... 73
6.73 Element <Software> ... 76
6.74 Element <SoftwareList> ... 77
6.75 Element <SourceFile> ... 77
6.76 Element <StudyVariable> ... 78
6.77 Element <StudyVariableList> ... 78
6.78 Element <StudyVariableQuantLayer> ... 79
6.79 Element <userParam> ... 80
7. Specific Comments on schema ... 80
7.1 File extension and compression ... 80
7.2 Referencing elements within the document ... 81
7.3 Unknown modifications ... 81
8. Conclusions ... 81
9. Authors and Contributors ... 81
10. References ... 82
11. Intellectual Property Statement ... 82
Copyright Notice ... 82

1. Introduction

1.1 Background

This document addresses the systematic description of quantifying molecules by mass spectrometry. A large number of different software packages are available that produce output in a variety of different formats. It is intended that mzQuantML will provide a single common format for software to represent, import or export quantitation values derived from mass spectrometry. These values typically report on peptides or proteins in the context of proteomics investigations but it is noted that similar structures are required in metabolomics, and, as such, structures have been developed that can capture small molecules descriptions and quantitative values.

mzQuantML has been developed with a view to supporting the following general tasks (more specific use cases are provided in Section 2):

T1. *The discovery of relevant results,* so that, for example, data sets in a database that use a particular technique or combination of techniques can be identified and studied by experimentalists during experiment design or data analysis.
T2. *The sharing of best practice*, so that, for example, analyses that have been particularly successful at quantifying a certain group of peptides/proteins can be interpreted by consumers of the data.
T3. *The evaluation of results*, so that, for example, sufficient information is provided about how a particular analysis was performed to allow the results to be critically evaluated.
T4. *The sharing of data sets,* so that, for example, public repositories can import or export data, or multi-site projects can share results to support integrated analysis.
T5. *The creation of a format for input to analysis software*, for example, allowing software to be designed that provides statistical significance on top of protein quantitation values.
T6. *An internal format for pipeline analysis software*, for example, allowing analysis software to store intermediate results from different stages of a quantitation pipeline, prior to the final results being assembled in a single mzQuantML file.

The description of the analysis of proteomics mass spectra requires that parts of the schema describe: (i) the identity and configuration of software used to perform the analysis and the protocol used to apply this software to the analysis; (ii) the quantitative data associated with molecules; and (iii) the way in which these relate to other techniques to form a proteomics workflow. Most of this document is concerned with (i) and (ii) – the identification of the key features of different techniques that are required to support the tasks T1 to T5 above. Models of type (iii) are created by developments in the context of the Functional Genomics Experimental Object Model (FuGE), which defines model components of relevance to a wide range of experimental techniques. Several components from FuGE are re-used in the development of mzQuantML.

This document presents a specification, not a tutorial. As such, the presentation of technical details is deliberately direct. The role of the text is to describe the model and justify design decisions made. The document does not discuss how the models should be used in practice, consider tool support for data capture or storage, or provide comprehensive examples of the models in use, beyond the key examples developed as part of the version 1.0 release. It is anticipated that tutorial material will be developed when the specification is stable.

1.2 Document Structure

The remainder of this document is structured as follows. Section 2 lists use cases mzQuantML is designed to support. Section 3 describes the terminology used. Section 4 describes how the specification presented in Section 6 relates to other specifications, both those that it extends and those that it is intended to complement. Section 5 discusses the reasoning behind several design decisions taken. Section 6 contains the documentation for the XML schema which is generated automatically and several parts of the schema are documented in more detail in Section 0. Conclusions are presented in Section 8.

2. Use Cases for mzQuantML

The development of mzQuantML is driven by some general principles, specific use cases and the goal of supporting specific techniques, as listed below. These were discussed and agreed at the development meeting in Tübingen in July 2011.

General principles, the format SHOULD support:

- Journal requirements for the reporting of quantitative proteomic data from mass spectrometry.
- Reporting according to the MIAPE Quant document.
- Submission of quantitative data to public databases.
- Data exchange between software tools, where data are defined as values about features (defined here as regions on MS1 mass spectra that report on a single peptide or small molecule), feature matches across different spectra or within spectra, peptides, proteins and protein groups.
- Import of data into statistical processing tools.
- The ability to reprocess or recreate the analysis workflow using the same parameters, assuming no manual steps have taken place.

Use cases, the format SHOULD capture:

- Final abundance values (relative or absolute) for peptides, proteins and protein groups where protein inference cannot be performed in an unambiguous manner.
- Quantitation values about peptide/protein modifications, such as post-translational modifications.
- Abundance values at the level of a single run (called an assay in this context) and logical groupings of runs (called study variables in this context), for which the user, for example, wishes to report relative values.
- The evidence trail for how final abundance values were calculated, such as the features used for quantifying peptides and proteins.
- Relationships between features either on different regions of the same MS run or on different MS runs that report on the same peptide or small molecule. These are particularly required for relative quantitation approaches.
- Details about pre-fractionation sufficient to describe the combination of multiple input data files (e.g. raw files) into a single assay where this has been performed.

The format SHOULD support the following specific techniques used in proteomics (see section 5.4 for examples of their encoding):

- MS1 label-free intensity
- MS1 label-based e.g. SILAC and metabolic labelling such as ^{15}N
- MS2 tag-based e.g. iTRAQ / TMT
- MS2 spectral counting

We expect that the format MAY also be able to cover the following techniques adequately, although these have not been tested in great detail at this stage, and we encourage further input from users of these techniques:

- Quantitation by selected reaction monitoring (SRM)
- Absolute quantitation based on averaging the intensities of features e.g. Waters Hi3 technique
- Small molecule quantitation (in metabolomics)
- MS2 intensity-based approaches
- MS2 label-based approaches

It is acknowledged that SRM in particular is an important quantitative technique in proteomics and, as such, we expect that the specifications for encoding SRM in mzQuantML will follow shortly after the publication of this specification, for example via a PSI Informational Document or Appendix to this specification.

3. Concepts and Terminology

This document assumes familiarity with XML Schema notation (www.w3.org/XML/Schema). The key words "MUST," "MUST NOT," "REQUIRED," "SHALL," "SHALL NOT," "SHOULD," "SHOULD NOT," "RECOMMENDED," "MAY," and "OPTIONAL" are to be interpreted as described in RFC-2119 [RFC2119].

4. Relationship to Other Specifications

The specification described in this document is not being developed in isolation; indeed, it is designed to be complementary to, and thus used in conjunction with, several existing and emerging models. Related specifications include the following:

1. *MIAPE Quant* (http://psidev.info/miape-quant). The Minimum Information About a Proteomics Experiment: Mass Spectrometry Quantification (MIAPE-Quant) document defines a checklist of information that should be reported about a quantitative proteomics study. It is expected that mzQuantML will be used to support MIAPE-Quant compliant submissions to public repositories.
2. *FuGE* (http://fuge.sourceforge.net). FuGE is a data model in UML, and an associated XML rendering, that represents various high-level concepts that are characteristic of functional genomics, such as investigations and protocols. FuGE has been developed by representatives of several standards bodies, with a view to making the representation of functional genomic data sets more consistent, and as such more easily shared and compared. The FuGE specifications are available from [Jones 07].

3. *mzML* (http://www.psidev.info/mzml/). mzML is the PSI standard for capturing mass spectra / peak lists resulting from mass spectrometry in proteomics. It is RECOMMENDED that mzQuantML should be used in conjunction with mzML, although it will be possible to use mzQuantML with other formats of mass spectra. This document does not assume familiarity with mzML.
4. *mzIdentML (*http://www.psidev.info/mzidentml/*).* mzIdentML is the PSI standard for peptide and protein identifications. It is RECOMMENDED that mzQuantML should be used in conjunction with mzIdentML, although it will be possible to use mzQuantML without a separate document storing identification evidence data.

4.1 Important concepts from FuGE

mzQuantML makes use of several components from FuGE to allow the format to be more easily integrated with other FuGE-based formats. However, FuGE is a large, flexible specification that can cover a variety of concepts not required for mzQuantML. In this release, various concepts from FuGE have been directly incorporated into the schema. Additional knowledge of FuGE is thus not required beyond this specification document.

4.2 The PSI Mass Spectrometry Controlled Vocabulary (CV)

The PSI-MS controlled vocabulary is intended to provide terms for annotation of mzML and mzQuantML files. The CV has been generated by collection of terms from software vendors and academic groups working in the area of mass spectrometry and proteome informatics. Some terms describe attributes that must be coupled with a numerical value attribute in the <cvParam> element (e.g. MS:1001870, “p-value for peptides ”) and optionally a unit for that value (e.g. MS:1001117, “theoretical mass”, units = dalton). The terms that require a value are denoted by having a “datatype” key-value pair in the CV itself: MS:1001172 " Mascot:expectation value" value-type:xsd:double. Terms that need to be qualified with units are denoted by have a “has_units” key in the CV itself (relationship: has_units: UO:0000221 ! dalton). The details of which terms are allowed or required in a given schema section is reported in the mapping file (Section 4.3).

As recommended by the PSI CV guidelines, psi-ms.obo should be dynamically maintained via the psidev-ms-vocab@lists.sourceforge.net mailing list that allows any user to request new terms in agreement with the community involved. Once a consensus is reached among the community the new terms are added within a few business days. If there is no obvious consensus, the CV coordinators committee should vote and make a decision. A new psi-ms.obo should then be released by updating the file on the CVS server without changing the name of the file (this would alter the propagation of the file to the OBO website and to other ontology services that rely on file stable URI). For this reason an internal version number with two decimals (x.y.z) should be increased:

- x should be increased when a first level term is renamed, added, deleted or rearranged in the structure. Such rearrangement will be rare and is very likely to have repercussion on the mapping.
- y should be increased when any other term except the first level one is altered.
- z should be increased when there is no term addition or deletion but just editing on the definitions or other minor changes.

The following ontologies or controlled vocabularies specified below may also be suitable or required in certain instances:

- Unit Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=unit)
- ChEBI (http://www.ebi.ac.uk/chebi/)
- OBI (Ontology of Biological Investigations - http://obi.sourceforge.net/)
- PSI Protein modifications workgroup - http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/mod/data/PSI-MOD.obo
- Unimod modifications database - http://www.unimod.org/obo/unimod.obo

4.3 Validation of controlled vocabulary terms

The correct usage of controlled vocabulary terms within mzQuantML is governed by the use of a mapping file which defines each XML location (XPath) where a <cvParam> instance can be used, and the allowed terms from the PSI-MS, or other, controlled vocabularies. The mapping file is read and interpreted by validation software, checking that the data annotation is consistent. The mapping file needs to be checked and updated when the structure of CV is changed, and in some instances when new terms are added to the CV. The specifications for the mapping file can be found here: http://www.psidev.info/validator. XML paths are associated with CV terms along with a requirement level (MAY, SHOULD or MUST) defining what should be reported by validation software if one of the mapped terms is not provided in an instance document. Example validation software based on the mapping file is available from http://code.google.com/p/mzquantml-validator/downloads/list.

4.3.1 Validation of values used within CvParams

An important decision has been made with mzQuantML that implementers should be aware of, which is different to standard PSI practice. The PSI-MS CV can contain a specification for the data type of a value that should be supplied in the value slot of the CvParam element.

```
[Term]
id: MS:1001171
name: Mascot:score
def: "The Mascot result 'Score'." [PSI:PI]
xref: value-type:xsd\:double "The allowed value-type for this CV term."
is_a: MS:1001116 ! single protein result details
is_a: MS:1001143 ! search engine specific score for peptides
is_a: MS:1001153 ! search engine specific score
```

In the example, the presence of the xref implies that if the CV term is used within a format, a value is to be provided according to the xsd:double data type. As an example:

```
<cvParam accession="MS:1001171" name="mascot:score" cvRef="PSI-MS" value="13.49"/>
```

In other PSI standards, the presence of the datatype xref within a CV term has been taken to mean: a value MUST be provided otherwise the cvParam element is deemed invalid. In the context of mzQuantML, data values are typically provided in <QuantLayer> elements, rather than within <cvParam>. As such, the data type specified by xref is implemented as meaning in mzQuantML:

If a value is provided within a <cvParam> element, it MUST follow the RECOMMENDED data type, but the presence of a value within <cvParam> is OPTIONAL in all cases. Validator implementations SHOULD NOT take the presence of the data type xref in the CV to imply that a value must be provided.

As an example, the following is valid mzQuantML:

```
<GlobalQuantLayer id="Pep_GQL1">
  <ColumnDefinition>
    <Column index="0">
     <DataType>
      <cvParam accession="MS:1001171" cvRef="PSI-MS" name="Mascot:score"/>
     </DataType>
    </Column>
   </ColumnDefinition>
   <DataMatrix>
    <Row object_ref="pep_GAPEIDVLEGETDTK_2_21711">83.67</Row>
    <Row object_ref="pep_QSTTFADCPVVPADPDILLAK_2_48178">52.13</Row>
    ...
```

It is recognized that this situation is not optimal. A preferable scenario would be to state in the CV mapping file whether the values should be provided or not. This will be a topic for future cross-PSI discussions and further development of the PSI validator framework.

5. Resolved Design and scope issues

There were several issues regarding the design of the format that were not clear cut, and a design choice was made that was not completely agreeable to everyone. So that these issues do not keep coming up, we document the issues here and why the decision that has been implemented was made.

5.1.1 Handling updates to the controlled vocabulary

There is a difficult issue with respect to how software should encode CV terms, such that changes to core can be accommodated. This issue is discussed at length in the mzML specification document [Martens11], and mzQuantML follows the same convention. In brief, when a new term is required, the file producers must contact the CV working group (via psidev-ms-vocab@lists.sourceforge.net) and request the new term. It is anticipated that problems may arise if a consumer of the file encounters a new CV term and they are not working from the latest version of the CV file. It has been decided that rather than aim for a workaround to this issue, it can be expected that data file consumers must ensure that the OBO file is up-to-date.

5.1.2 Use of mzQuantML for analysis pipelines

The primary use case driving the development of mzQuantML is to allow reporting of quantitation values and to allow data to be shared between different tools. It is clear that the proteome bioinformatics community would benefit from a format that can be used internally for capturing the associations between features, peptides and proteins. As such, there is limited support for using mzQuantML for software pipelines, for example by allowing multiple <PeptideConsensusList> elements, which can be mapped as inputs and outputs of a software pipeline using <DataProcessingList>. A mandatory Boolean attribute is provided on <PeptideConsensusList> called finalResult, which allows a consumer of the data file to know which list of peptide results is considered the final result, for example for loading into a database where intermediate results are not required.

5.2 Encoding zeroes, nulls, infinity and calculation errors

In various parts of the schema, most notably in <DataMatrix> within QuantLayers, numerical values must be provided. In data analysis, it is possible that different types of value arise: null values where an entity has not been measured, a zero value where the entity has been measured but has zero value, infinity values for example in ratios where the denominator is zero, or calculation errors resulting in the common “not a number” (NaN type). In some cases, there is ambiguity with respect to these cases: e.g. in spectral counting if no peptide-spectrum matches are observed for a given protein, it is open for debate as to whether its abundance is zero or null.

In the XML Schema specifications under the xsd:double type, there is a clear mapping for some of these concepts but not all. As such, rows within a <DataMatrix> can contain a list of a union data type between xsd:double and the token “null” to allow nulls to be adequately captured. The encodings are summarized in Table 1.

Value type	mzQuantML encoding	Notes
Zero	0.0	For intensity-based approaches, where a feature has been measured but there is no intensity, the “0.0” value MUST be given.
infinity	INF	If ratios are included and the denominator is zero, the “INF” value MUST be given.
Null	null	If an entity has not been measured, for example in an

		<AssayQuantLayer> where a certain peptide was found for some but not all assays, the “null” value MUST be given. A zero value is not appropriate since, in this example, the peptide may have been ionised but was not measured in one or more assays. Programming language bindings for mzQuantML MUST handle this explicitly, since “null” is technically encoded in mzQuantML as an xsd:token, and automated data type bindings will fail.
Result of calculation error	NaN	If a calculation error has occurred that does not result in a true null, INF or zero value, the value “NaN” must be given.

Table 1 The RECOMMENDED methods for encoding zero values, "not a number" (NaN), nulls and infinity values in mzQuantML.

5.3 Protein grouping

Software exporting results into mzQuantML MAY include results at the protein level or protein group level, depending on method or analysis software used. However, exporters SHOULD NOT include quant layers at both levels. It is expected that an import utility, such as a public database will import results at the protein group level in preference to results at the protein level if a file contains both types of quant layer.

The CV mapping file contain a rule indicating that terms SHOULD be attached to the referenced proteins within a <ProteinGroup> describing their role within the group, such as the arbitrary group representative (currently “anchor protein”), “sequence same-set” and so on. In tandem with efforts to improve protein grouping representation in mzIdentML, these terms may be subject to change through future PSI Informational Documents. It is likely that a rule will be introduced ensuring that exactly one proteins within each group is flagged as the group representative and, as such, exporters are RECOMMENDED to include exactly one <ProteinRef> element within each <ProteinGroup> with the CV term “anchor protein” (MS:1001591).

5.4 Comments on Specific Use Cases

Several use cases for mzQuantML were considered during its development. Each of these use cases has a corresponding example file that exercises the relevant part of the schema and provides a reference implementation example (see supporting documentation). Authors of software that create mzQuantML are encouraged to examine the examples that accompany this format release before implementing the writer. Further, such authors are encouraged to use the validator (Section 5.5) before releasing any new mzQuantML export code and working with the PSI PI Working Group to resolve any issues. In the subsections below, we describe the use cases supported in version 1.0.

In the version 1.0 release of mzQuantML, four techniques are included i) MS1 label-free intensity, ii) MS1 label-based (e.g. SILAC), iii) MS2 tag-based (e.g. iTRAQ / TMT) and iv) spectral counting label free. Exactly one of four specific CV terms MUST be included in the <AnalysisSummary> element, allowing the validation software to determine which semantic rule set to apply:

```
<cvParam accession="MS:1002018" cvRef="PSI-MS" name="MS1 label-based analysis"/>
<cvParam accession="MS:1001834" cvRef="PSI-MS" name="LC-MS label-free quantitation analysis"/>
<cvParam accession="MS:1002023" cvRef="PSI-MS" name="MS2 tag-based analysis"/>
<cvParam accession="MS:1001836" cvRef="PSI-MS" name="spectral counting quantitation analysis"/>
```

As such, if a software package has employed more than one technique in the same analysis, these MUST be encoded in separate files. Additional techniques will be supported in the near future, through the release of updated semantic validation rules.

5.4.1 MS1 label-free intensity

Certain software packages (such as Progenesis LC-MS), align parallel runs in the retention time axis, and then quantify the identical feature region in all runs, even if there is no intensity in that region (thus giving a zero value). As such, the data can be represented in a regular fashion with a <PeptideConsensus> element referencing exactly one feature for each assay. If the same peptide occurs in a different charge state or with a different modification, this is modelled by a different <PeptideConsensus> element. No attempt is made in the <PeptideConsensusList> to model the summed abundance of different features within the same assay that report on the same peptide. Other packages that, for example, only quantify the peptides that have been confidently identified SHOULD use the same encoding, reporting the "null" datatype for any peptides not identified in particular assays.

QuantLayers SHOULD be provided on the <PeptideConsensusList> and the <ProteinGroupList> or <ProteinList>. QuantLayers MAY be provided on the <FeatureList> for reporting additional data types about the features calculated prior to the feature matching process. However, adding QuantLayers to the <FeatureList> can lead to verbose files, so this method of encoding is NOT RECOMMENDED.

Example files are provided from export of the same MS data analysed with both Progenesis LC-MS and MaxQuant at http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/label-free/

5.4.2 MS1 label-based

In MS1 label-based approaches two (or more) samples are mixed and analysed once by MS. In many approaches, pairs of features separated by a predictable mass shift are identified that report the relative abundance of the same peptide, from which a ratio can be calculated. To illustrate how mzQuantML encodes such an approach, the following text describes the encoding of a SILAC approach with a +8 shift for Lys and +10 shift for Arg, analysed in 3 replicates runs.

See examples: http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/MS1Label/

For each replicate, there SHOULD be two <Assay> elements, each of which MUST refer to the same raw file(s) in <RawFileGroup>. One of the <Assay> elements MUST contain details of the modification used to differentiate the peptide, such as the mass shift for the heavy lys/arg. The other <Assay> must include the CV term "unlabeled sample" (MS:1002038).

If the data exporter wishes to communicate the full evidence trail, the primary results from analysis of each <RawFileGroup> SHOULD be represented as a <FeatureList>. The <FeatureList> MAY contain data types about these features, including their raw intensity value. The next stage of the analysis is the finding of pairs of features that report on the same peptide (one of which has the expected mass shift). The result of this analysis for the three replicates SHOULD be captured by one or more <PeptideConsensusList> elements.

The exporter MAY encode results matched across replicate runs in one single <PeptideConsensusList> or maintain separate lists for each replicate run, however, it is REQUIRED that there is exactly one <PeptideConsensusList> with finalResult = "true", if the file could be imported by a public database. Within each <PeptideConsensusList>, there SHOULD be QuantLayers that report the abundance of peptides as calculated for individual assays or study variables as appropriate, and the exporter MAY encode these using a <RatioQuantLayer>. As with other approaches, if the identity of a peptide is unknown, the <PeptideConsensus> element MUST NOT have the <PeptideSequence> sub-element, but quantitative values MAY still be provided in the associated QuantLayers.

The schema only allows one <ProteinGroupList> and/or one <ProteinList> and as such, if the exporter wishes to communicate values about proteins, the instance MUST contain one <ProteinGroupList> and/or one <ProteinList>, including the relevant QuantLayers for assays or study variables.

5.4.3 MS2 spectral counting

MS2 spectral counting approaches use the number of peptide-spectrum matches assigned to a given protein to estimate the protein abundance in the sample, following a variety of different normalisation schemes. As such, there is no requirement to model the intensity of MS1 features or provide QuantLayers on the <PeptideConsensusList>. A <PeptideConsensusList> MAY be provided to report the peptide sequences identified, although if accompanying mzIdentML files exist, it is not REQUIRED. QuantLayers SHOULD only be provided on the <ProteinGroupList> and/or <ProteinList> to capture the abundance of proteins within assays or study variables.

See examples:
http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/spectral-count/

5.4.4 MS2 tag-based

In MS2 tag-based approaches multiple samples are initially prepared in parallel and each set of peptides is modified by a chemical tag (such as an isobaric tag used in iTRAQ) by which it can be differentiated in MS2. The samples are mixed and analysed once by MS/MS, using the relative intensities of the tags when measured in MS2 to calculate the relative intensities of the tagged peptide in each of the source samples.

In this example, the use of iTRAQ with tags 114, 115, 116 and 117 Daltons in three replicates is described. The primary results from analysis of each <RawFileGroup> SHOULD be represented as a <FeatureList>, as such producing three lists, one for each replicate. Each <FeatureList> SHOULD contain at least one quant layer, of type <MS2AssayQuantLayer> or <MS2RatioQuantLayer> (where the ratios captured describe all assays e.g. 117/114, 116/114, 115/114) depending on what the data exporter wishes to communicate. The exporter MAY also include an <MS2StudyVariableQuantLayer>. It is REQUIRED that each <Assay> element describes the iTRAQ tag contained, and as such, in an <MS2AssayQuantLayer> or <MS2RatioQuantLayer> a data consumer could work out the correspondence between an intensity value given for an assay (or ratio of assays) and the source intensity (or ratio of intensity) of the tag in the MS2 spectrum.

In many MS2 tagging approaches, quantitative values can be obtained for the same peptide in multiple different MS2 scans. In this instance, the exporter SHOULD represent the raw intensity values obtained from each scan within the <MS2AssayQuantLayer> under <FeatureList> and the final intensity values for peptides (summed or averaged over the different scans) within <AssayQuantLayer> under <PeptideConsensusList>. See example file: http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/MS2Tag/iTraq_4plex_example_from_xTracker.mzq where this approach has been followed.

The concept of a <Feature> in mzQuantML describes a region on an MS1 spectrum/spectra and as such elements of <Feature> SHOULD be created to capture only the m/z and charge of the parent ion from which the iTRAQ intensities are calculated. It is noted that in some cases, the retention time of the MS1 features are not known in iTRAQ yet the RT attribute is mandatory on <Feature>. In this case, the “null” value should be used, as discussed in section 5.2.

The exporter MAY encode results matched across replicate runs in one single <PeptideConsensusList> or maintain separate lists for each replicate run, however, it is REQUIRED that there is exactly one <PeptideConsensusList> with finalResult = “true”, if the file may be imported to a public database. Within each

<PeptideConsensusList>, there SHOULD be QuantLayers that report the abundance of peptides as calculated for individual assays or study variables as appropriate, and the exporter MAY encode these using a <RatioQuantLayer>. As with other approaches, if the identity of a peptide is unknown, the <PeptideConsensus> element MUST NOT have the <PeptideSequence> sub-element, but quantitative values MAY still be provided in the associated QuantLayers.

Examples: http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/MS2Tag/

5.5 Semantic validation rules

In tandem with the CV mapping file, a set of additional semantic validation rules have been written in natural language, and encoded in the mzQuantML validation software, to ensure that specific rules are followed for the four types of technique that are included in the version 1.0 release (Section 5.4).

The semantic validation rules are given here: http://code.google.com/p/mzquantml/source/browse/trunk/schema/ and validation software is available from here: http://code.google.com/p/mzquantml-validator/.

The use of semantic validation rules to place additional constraints on the core schema, means that support for additional techniques, such as SRM can be added to mzQuantML without requiring updates to the core schema. We anticipate that this will standardised through the release of a new Appendix to this document or a PSI Informational document.

5.6 Other supporting materials

The following example instance documents are available and between them cover all the techniques and use cases supported, including working examples for techniques not formally supported at this time (Section 2).

All example files can be downloaded manually from:
http://code.google.com/p/mzquantml/source/browse/trunk/examples/version1.0-rc3/

5.7 Open Issues

5.7.1 Support for metabolomics

During the development process for mzQuantML, it has been noted that while this is primarily developed as a format for representing peptide and protein quantitative values, similar structures are required for representing small molecules for metabolomic studies. As such, an extension has been created to the schema that attempts to model small molecules, although it should be noted that at this stage, this has not been tested to the same extent as other parts of the model. The decision was taken to include the core structures in the version 1.0 release (but no semantic validation rules), thus allowing groups working in this field to test a release version of the format, although no guarantee is made that this structure is yet optimal. We encourage further input from researchers working in the small molecule field.

6. Model in XML Schema

An overview of the schema is presented in Figure 1. The following documentation is automatically generated from the XML Schema.

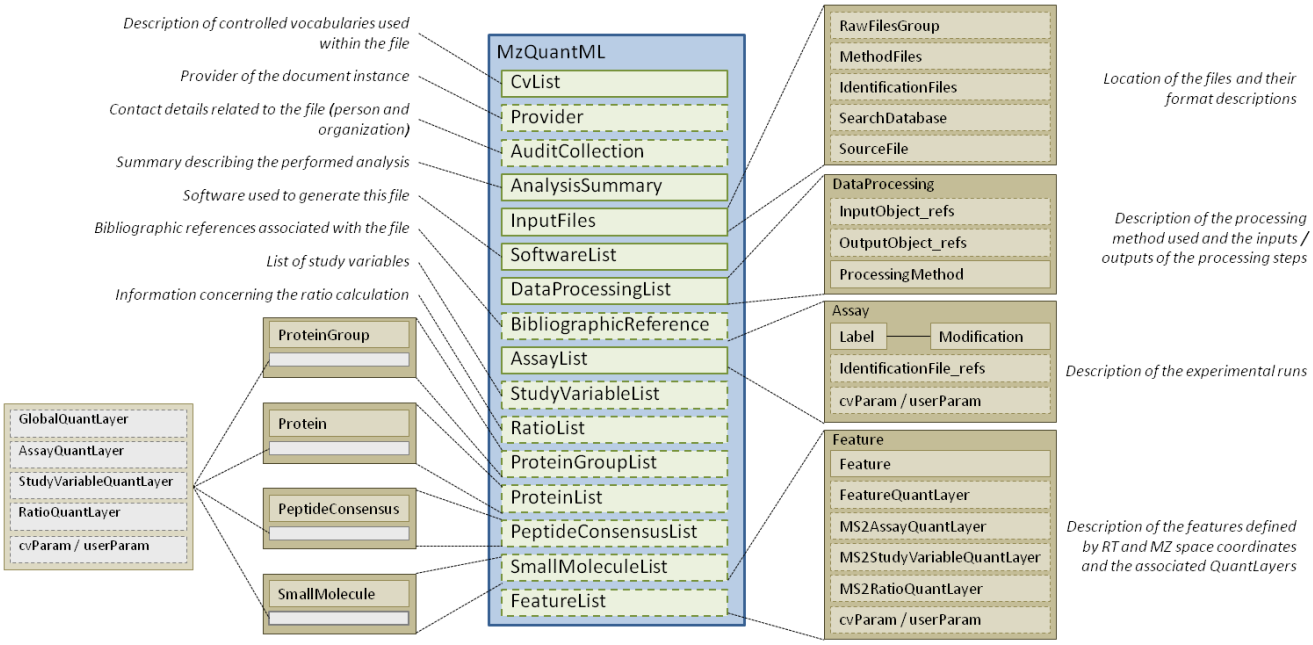


Figure 1 A diagrammatic overview of the mzQuantML schema.

6.1 Element <MzQuantML>

Definition: Root element of the instance document.

Type: MzQuantMLType

Attributes:

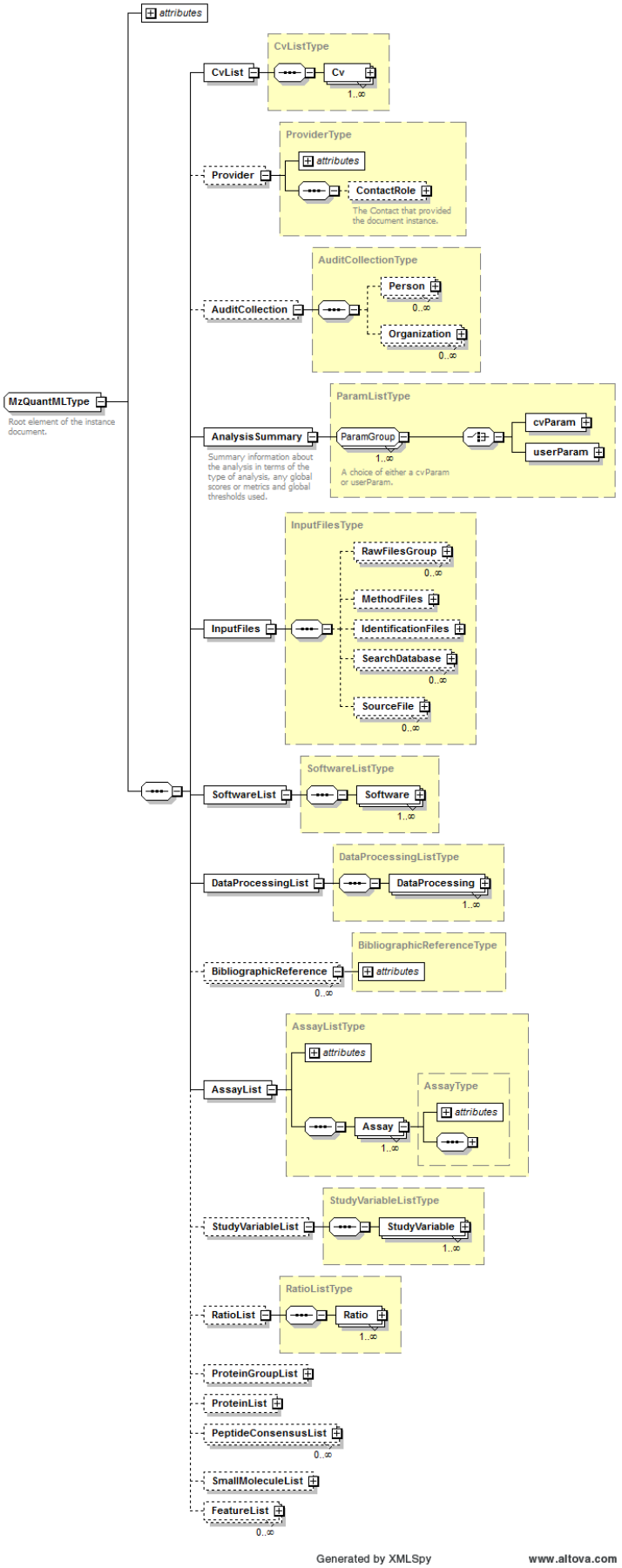
Attribute Name	Data Type	Use	Definition
creationDate	xsd:dateTime	optional	The date on which the file was produced.
id	xsd:ID	required	The unique identifier for the mzQuantML file or experiment.
name	xsd:string	optional	Optional name for the file or experiment.
version	versionRegex	required	The version of the schema this instance document refers to, in the format x.y.z. Changes to z should not affect prevent instance documents from validating.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
CvList	1	1	The list of controlled vocabularies used in the file.
Provider	0	1	The provider of the document in terms of the Contact and the software the produced the document instance.
AuditCollection	0	1	The complete set of Contacts (people and organisations) for this file.
AnalysisSummary	1	1	Summary information about the analysis in terms of the type of analysis, any global scores or metrics and global thresholds

			used.
InputFiles	1	1	All the raw files, identification files and databases used in the quantitation.
SoftwareList	1	1	List and descriptions of software used to acquire and/or process the data in this file.
DataProcessingList	1	1	List and descriptions of data processing applied to this data.
BibliographicReference	0	unbounded	Represents bibliographic references.
AssayList	1	1	The list of assays represented by the file, where each assay captures the concept of one sample analysed - this can be one or more raw files.
StudyVariableList	0	1	The list of experimental conditions used to group results.
RatioList	0	1	The definition of ratios of study variables or assays, referenced elsewhere in the document.
ProteinGroupList	0	1	The list of all groups of proteins with conflicting evidence for which quantitation values are being reported along with quantitative values about those protein groups. If quantitation is done on individual proteins only, ProteinGroupsList should not be included.
ProteinList	0	1	The list of all individual proteins (i.e. ungrouped) for which quantitation values are being reported. If quantitation is done on protein groups, the constituent proteins should be listed here with no QuantLayers.
PeptideConsensusList	0	unbounded	The list of all peptides objects for which quantitation values are reported.
SmallMoleculeList	0	1	List of small molecules and associated data values.
FeatureList	0	unbounded	All the features identified on a single raw file or raw file group.

Graphical Context:



Example Context:

```
<MzQuantML creationDate="2012-11-30T00:00:00.000+01:00" version="1.0.0" id="MZQ_SPC_EX1"
xsi:schemaLocation="http://psidev.info/psi/pi/mzQuantML/1.0.0-rc3 ../../../schema/mzQuantML_1_0_0-
rc3.xsd" xmlns="http://psidev.info/psi/pi/mzQuantML/1.0.0-rc3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <CvList>
        <Cv fullName="Proteomics Standards Initiative Protein Modifications" version="1.010.7"
uri="http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/mod/data/PSI-MOD.obo" id="MOD"/>
        <Cv fullName="Proteomics Standards Initiative Mass Spectrometry Vocabulary" version="3.41.0"
uri="http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo"
id="MS"/>
        <Cv fullName="UNIMOD CV for modifications" version=""
uri="http://www.unimod.org/obo/unimod.obo" id="UNIMOD"/>
        <Cv fullName="Unit Ontology" version=""
uri="http://obo.cvs.sourceforge.net/viewvc/obo/obo/ontology/phenotype/unit.obo" id="UO"/>
    </CvList>
  ...
</MzQuantML>
```

6.2 Element <Affiliation>

Definition: The organization a person belongs to.

Type: AffiliationType

Attributes:

Attribute Name	Data Type	Use	Definition
organization_ref	xsd:IDREF	required	A reference to the organization this contact belongs to.

Subelements: none

Example Context:

```
<Affiliation organization_ref="ORG_UOL"/>
```

6.3 Element <AnalysisSummary>

Definition: Summary information about the analysis in terms of the type of analysis, any global scores or metrics and global thresholds used.

Type: ParamListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	1	unbounded	A single user-defined parameter.

Example Context:

```
  <AnalysisSummary>
    <cvParam cvRef="PSI-MS" accession="MS:1002018" name="MS1 label-based analysis"/>
    <cvParam cvRef="PSI-MS" accession="MS:1001835" name="SILAC quantitation analysis"/>
    <cvParam cvRef="PSI-MS" accession="MS:1002001" name="MS1 label-based raw feature
quantitation" value="true"/>
    <cvParam cvRef="PSI-MS" accession="MS:1002002" name="MS1 label-based peptide level
quantitation" value="true"/>
    <cvParam cvRef="PSI-MS" accession="MS:1002003" name="MS1 label-based protein level
quantitation" value="false"/>
    <cvParam cvRef="PSI-MS" accession="MS:1002004" name="MS1 label-based proteingroup level
quantitation" value="false"/>
  ...
</AnalysisSummary>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AnalysisSummary
MUST supply a *child* term of MS:1001833 (quantitation analysis summary) only once
  e.g.: MS:1001835 (SILAC quantitation analysis)
```

e.g.: MS:1001837 (iTRAQ quantitation analysis)
e.g.: MS:1001838 (SRM quantitation analysis)
e.g.: MS:1001839 (metabolic labeling 14N / 15N quantitation analysis)
e.g.: MS:1002001 (MS1 label-based raw feature quantitation)
e.g.: MS:1002002 (MS1 label-based peptide level quantitation)
e.g.: MS:1002003 (MS1 label-based protein level quantitation)
e.g.: MS:1002004 (MS1 label-based proteingroup level quantitation)
e.g.: MS:1002010 (TMT quantitation analysis)
e.g.: MS:1002015 (spectral count peptide level quantitation)
et al.

Example cvParams:

```
<cvParam accession="MS:1001834" cvRef="PSI-MS" name="LC-MS label-free quantitation analysis"/>
<cvParam accession="MS:1002019" cvRef="PSI-MS" value="false" name="label-free raw feature quantitation"/>
<cvParam accession="MS:1002020" cvRef="PSI-MS" value="true" name="label-free peptide level quantitation"/>
<cvParam accession="MS:1002021" cvRef="PSI-MS" value="true" name="label-free protein level quantitation"/>
<cvParam accession="MS:1002022" cvRef="PSI-MS" value="false" name="label-free proteingroup level quantitation"/>
<cvParam accession="MS:1002018" cvRef="PSI-MS" name="MS1 label-based analysis"/>
<cvParam accession="MS:1002001" cvRef="PSI-MS" value="true" name="MS1 label-based raw feature quantitation"/>
<cvParam accession="MS:1002002" cvRef="PSI-MS" value="true" name="MS1 label-based peptide level quantitation"/>
<cvParam accession="MS:1002003" cvRef="PSI-MS" value="true" name="MS1 label-based protein level quantitation"/>
<cvParam accession="MS:1002004" cvRef="PSI-MS" value="false" name="MS1 label-based proteingroup level quantitation"/>
<cvParam cvRef="PSI-MS" accession="MS:1001835" name="SILAC quantitation analysis"/>
<cvParam accession="MS:1002023" cvRef="PSI-MS" name="MS2 tag-based analysis"/>
<cvParam accession="MS:1002024" cvRef="PSI-MS" value="true" name="MS2 tag-based feature level quantitation"/>
<cvParam accession="MS:1002025" cvRef="PSI-MS" value="true" name="MS2 tag-based peptide level quantitation"/>
<cvParam accession="MS:1002026" cvRef="PSI-MS" value="true" name="MS2 tag-based protein level quantitation"/>
<cvParam accession="MS:1002027" cvRef="PSI-MS" value="false" name="MS2 tag-based proteingroup level quantitation"/>
<cvParam accession="MS:1001836" cvRef="PSI-MS" name="spectral counting quantitation analysis"/>
<cvParam accession="MS:1002015" cvRef="PSI-MS" value="false" name="spectral count peptide level quantitation"/>
<cvParam accession="MS:1002016" cvRef="PSI-MS" value="true" name="spectral count protein level quantitation"/>
<cvParam accession="MS:1002017" cvRef="PSI-MS" value="false" name="spectral count proteingroup level quantitation"/>
```

6.4 Element <Assay>

Definition: Describes a single analysis of a sample (e.g. with the channel mapping in iTRAQ), which could constitute multiple raw files e.g. if pre-separation steps have occurred.

Type: AssayType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
name	xsd:string	optional	Human readable name for the assay.
rawFilesGroup_ref	xsd:IDREF	optional	A reference to the RawFilesGroup that the Assay is linked to.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Label	1	1	A specification of labels or tags used to define the assay within the raw file, such as heavy labelling or iTRAQ tag mass. The

			Label and Modification is mandatory so a specific term is provided under Modification for unlabeled sample for label-free and, for example, so-called light samples in a labelling experiment.
IdentificationFile_refs	0	1	One or more identification files used within this assay.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
    <Assay id="a_3151552639304580883" rawFilesGroup_ref="rfg_15787140749886983179">
      <Label>
        <Modification massDelta="8.0141988132" >
          <cvParam cvRef="PSI-MOD" accession="MOD:00582" name="6x(13)C,2x(15)N labeled L-lysine"
value="Lys8"/>
        </Modification>
        <Modification massDelta="10.0082686" >
          <cvParam cvRef="PSI-MOD" accession="MOD:00587" name="6x(13)C,4x(15)N labeled L-arginine"
value="Arg10"/>
  ...
</Assay>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AssayList/Assay
MAY supply a *child* term of MS:1002110 (assay attribute) one or more times
```

6.5 Element <AssayList>

Definition: The list of assays represented by the file, where each assay captures the concept of one sample analysed - this can be one or more raw files.

Type: AssayListType

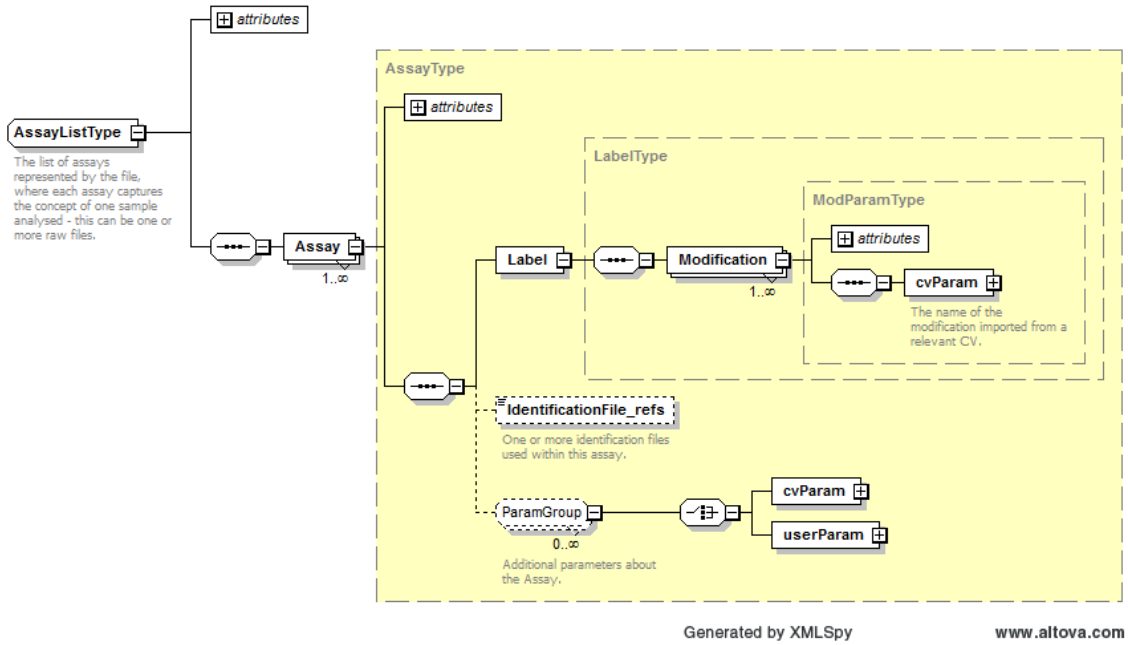
Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Assay	1	unbounded	Describes a single analysis of a sample (e.g. with the channel mapping in iTRAQ), which could constitute multiple raw files e.g. if pre-separation steps have occurred.

Graphical Context:



Example Context:

```
<AssayList id="AssayList1">
    <Assay rawFilesGroup_ref="raw1" name="114" id="_114">
        <Label>
            <Modification massDelta="145.0">
                <cvParam accession="MOD:01522" cvRef="PSI-MOD" value="114" name="iTRAQ4plex-114
reporter fragment"/>
            </Modification>
        </Label>
  ...
</AssayList>
```

6.6 Element <AssayQuantLayer>

Definition:

Depending on context:

1: Quant layer for reporting data values about protein groups related to different assays i.e. the column index MUST refer to Assays defined in the file.

2: Quant layer for reporting data values about proteins related to different assays i.e. the column index MUST refer to Assays defined in the file.

3: Quant layer for reporting data values about peptides related to different assays i.e. the column index MUST refer to Assays defined in the file.

4: Quant layer for reporting data values about small molecules related to different assays i.e. the column index MUST refer to Assays defined in the file.

Type: QuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataType	1	1	Type of data in the quant layer e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFoldChange|....".
ColumnI	1	1	Space separated unique identifiers for each column of data, MUST refer

ndex			to an object in the file i.e. StudyVariable or Assay, depending on the context where the QuantLayer resides.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

```
<AssayQuantLayer id="Pep_AQL1">
    <DataType>
        <cvParam accession="MS:1001891" cvRef="PSI-MS" name="Progenesis:peptide normalised abundance"/>
    </DataType>
    <ColumnIndex>ass_0 ass_1 ass_2 ass_3 ass_4 ass_5 ass_6 ass_7 ass_8 ass_9 ass_10 ass_11</ColumnIndex>
    <DataMatrix>
        <Row object_ref="pep_GAPEIDVLEGETDTK_2_21711">9079.67 6680.893 6869.919 7069.431 6872.01 7488.588
5045.095 7063.636 6636.347 3902.41 4320.357 4045.954</Row>
  ...
</AssayQuantLayer>
```

6.7 Element <Assay_refs>

Definition: Reference to the assays that are contained within this study variable.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context: `<Assay_refs>ASS_0 ASS_1 ASS_2 ASS_3 ASS_4</Assay_refs>`

6.8 Element <AuditCollection>

Definition: The complete set of Contacts (people and organisations) for this file.

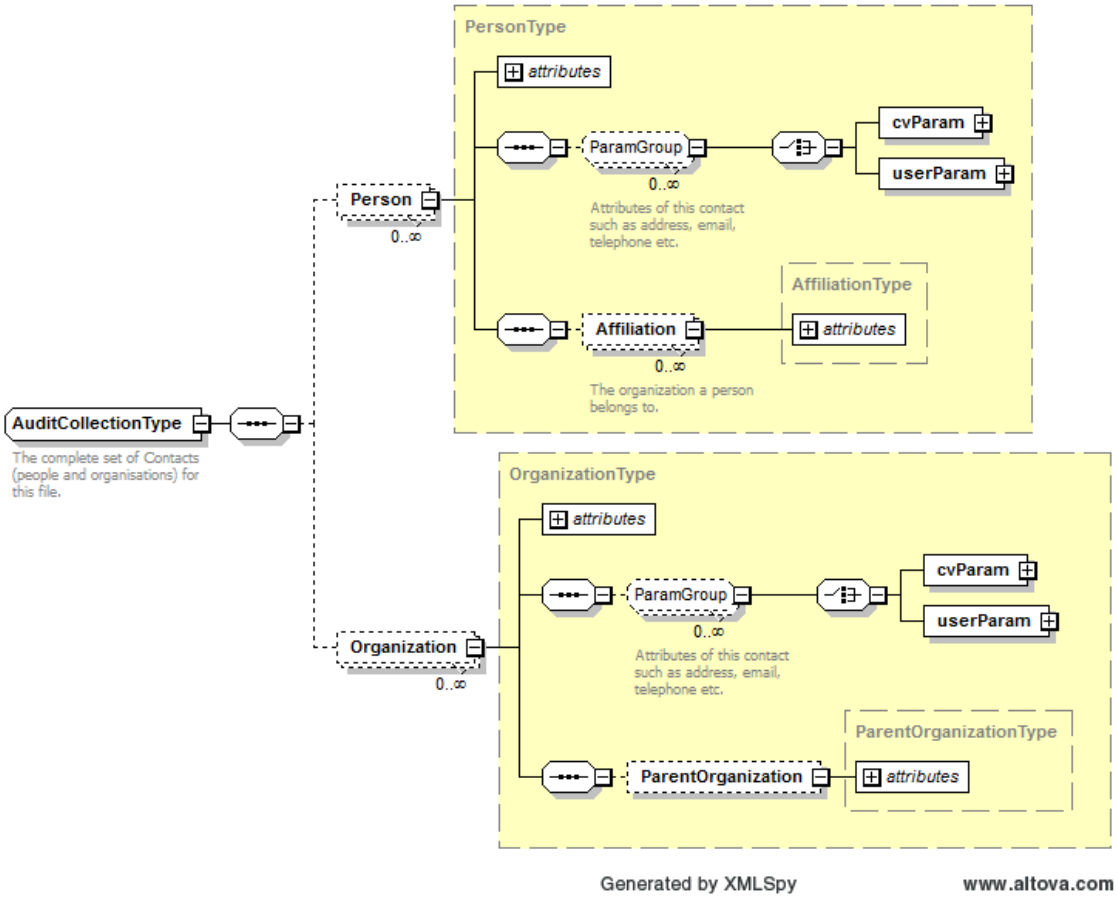
Type: AuditCollectionType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Person	0	unbounded	A person's name and contact details. Any additional information such as the address, contact email etc. should be supplied using CV parameters or user parameters.
Organization	0	unbounded	Organizations are entities like companies, universities, government agencies. Any additional information such as the address, email etc. should be supplied either as CV parameters or as user parameters.

Graphical Context:



Example Context:

```
<AuditCollection>
    <Person lastName="Mayer" firstName="Gerhard" id="PERS_GM" name="Gerhard Mayer">
        <Affiliation organization_ref="ORG_MPC"/>
    </Person>
    <Organization id="ORG_Univ_Bochum" name="University of Bochum"/>
    <Organization id="ORG_MPC" name="MPC - Medizinisches Proteom Center, Bochum">
        <ParentOrganization organization_ref="ORG_Univ_Bochum"/>
  ...
</AuditCollection>
```

6.9 Element <BibliographicReference>

Definition: Represents bibliographic references.

Type: BibliographicReferenceType

Attributes:

Attribute Name	Data Type	Use	Definition
authors	xsd:string	optional	The names of the authors of the reference.
doi	xsd:string	optional	The DOI of the referenced publication.
editor	xsd:string	optional	The editor(s) of the reference.
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
issue	xsd:string	optional	The issue name or number.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a

			human-readable name for the instance.
pages	xsd:string	optional	The page numbers.
publication	xsd:string	optional	The name of the journal, book etc.
publisher	xsd:string	optional	The publisher of the publication.
title	xsd:string	optional	The title of the BibliographicReference.
volume	xsd:string	optional	The volume name or number.
year	xsd:int	optional	The year of publication.

Subelements: none

Example Context:

```
<BibliographicReference doi="doi:10.1089/omi.2012.0042" title="A software toolkit and interface for performing stable isotope labelling and top3 quantification using Progenesis LC-MS" pages="489-495" issue="9" volume="16" year="2012" publication="OMICS: A Journal of Integrative Biology" authors="D. Qi, P. Brownridge, D. Xia, K. Mackay, F. F. Gonzalez-Galarza, J. Kenyani, V. Harman, R. J. Beynon and A. R. Jones" id="BF_DQ1"/>
```

6.10 Element <Column>

Definition: The datatype and index position of one column of data in the DataMatrix.

Type: ColumnType

Attributes:

Attribute Name	Data Type	Use	Definition
index	xsd:nonNegativeInteger	required	The column position within the data matrix - incrementing positive integers starting from zero.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataType	1	1	Type of data in the column e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance".

Example Context:

```
<Column index="0">
  <DataType>
    <cvParam cvRef="PSI-MS" accession="MS:1001141" name="intensity of precursor ion"/>
  </DataType>
 </Column>
<Column index="1">
```

6.11 Element <ColumnDefinition>

Definition: Definition of the data types in each column.

Type: ColumnDefinitionType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Column	1	unbounded	The datatype and index position of one column of data in the DataMatrix.

Example Context:

```
<ColumnDefinition>
  <Column index="0">
    <DataType>
      <cvParam cvRef="PSI-MS" accession="MS:1001141" name="intensity of precursor ion"/>
```

```
        </DataType>
      </Column>
      <Column index="1">
        <DataType>
  ...
</ColumnDefinition>
```

6.12 Element <ColumnIndex>

Definition: **Depending on context:**
1: Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. StudyVariable or Assay, depending on the context where the QuantLayer resides.
2: Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

```
<ColumnIndex>RATIO_vIa_30min_vs_vIb_120min RATIO_20a_30min_vs_20b_120min
RATIO_40a_30min_vs_40b_120min RATIO_60a_30min_vs_60b_120min RATIO_80a_30min_vs_80b_120min
RATIO_Ia_30min_vs_Ib_120min</ColumnIndex>
```

6.13 Element <ContactRole>

Definition: The role of the Contact that provided the document instance.

Type: ContactRoleType

Attributes:

Attribute Name	Data Type	Use	Definition
contact_ref	xsd:IDREF	required	When a ContactRole is used, it specifies which Contact the role is associated with.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Role	1	1	The roles (lab equipment sales, contractor, etc.) the Contact fills.

Example Context:

```
<ContactRole contact_ref="PERS_GM">
    <Role>
        <cvParam cvRef="MS" accession="MS:1001271" name="researcher"/>
    </Role>
</ContactRole>
```

6.14 Element <Cv>

Definition: A source controlled vocabulary from which cvParams will be obtained.

Type: CvType

Attributes:

Attribute Name	Data Type	Use	Definition
fullName	xsd:string	required	The full name of the CV.
id	xsd:ID	required	The unique identifier of this cv within the document to be referenced by cvParam elements.
uri	xsd:anyURI	required	The URI of the source CV.

version	xsd:string	optional	The version of the CV.

Subelements: none

Example Context:

```
<Cv id="PSI-MS" fullName="Proteomics Standards Initiative Mass Spectrometry Vocabularies" uri="http://psidev.cvs.sourceforge.net/viewvc/*checkout*/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo" version="3.40.0"/>
```

6.15 Element <CvList>

Definition: The list of controlled vocabularies used in the file.

Type: CvListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Cv	1	unbounded	A source controlled vocabulary from which cvParams will be obtained.

Example Context:

```
<CvList>
    <Cv fullName="Proteomics Standards Initiative Protein Modifications" version="1.010.7" uri="http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/mod/data/PSI-MOD.obo" id="MOD"/>
    <Cv fullName="Proteomics Standards Initiative Mass Spectrometry Vocabulary" version="3.41.0" uri="http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo" id="MS"/>
    <Cv fullName="UNIMOD CV for modifications" version="" uri="http://www.unimod.org/obo/unimod.obo" id="UNIMOD"/>
    <Cv fullName="Unit Ontology" version="" uri="http://obo.cvs.sourceforge.net/viewvc/obo/obo/ontology/phenotype/unit.obo" id="UO"/>
</CvList>
```

6.16 Element <cvParam>

Definition: A single entry from an ontology or a controlled vocabulary.

Type: CVParamType

Attributes:

Attribute Name	Data Type	Use	Definition
accession	xsd:string	required	The accession or ID number of this CV term in the source CV.
cvRef	xsd:IDREF	required	A reference to the cv element from which this term originates.
name	xsd:string	required	The name of the parameter.
unitAccession	xsd:string	optional	An accession number identifying the unit within the OBO foundry Unit CV.
unitCvRef	xsd:string	optional	If a unit term is referenced, this attribute MUST refer to the CV 'id' attribute defined in the cvList in this file.
unitName	xsd:string	optional	The name of the unit.
value	xsd:string	optional	The user-entered value of the parameter.

Subelements: none

Example Context:

```
<cvParam accession="MS:1001524" cvRef="PSI-MS" unitCvRef="UO" unitName="dalton" unitAccession="UO:0000221" value="63.998285" name="fragment neutral loss"/>
```

6.17 Element <DatabaseName>

Definition: The database name may be given as a cvParam if it maps exactly to one of the release databases listed in the CV, otherwise a userParam should be used.

Type: ParamType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.
userParam	1	1	A single user-defined parameter.

Example Context:

```
<DatabaseName>
  <userParam name="SwissProtHuman554TD_20080526.fasta"/>
</DatabaseName>
```

cvParam Mapping Rules:

```
Path /MzQuantML/InputFiles/SearchDatabase/DatabaseName
SHOULD supply a *child* term of MS:1001013 (database name) only once
  e.g.: MS:1001084 (database nr)
  e.g.: MS:1001104 (database UniProtKB/Swiss-Prot)
  e.g.: MS:1001142 (database IPI_human)
  e.g.: MS:1001285 (database IPI_mouse)
  e.g.: MS:1001286 (database IPI_rat)
  e.g.: MS:1001287 (database IPI_zebrafish)
  e.g.: MS:1001288 (database IPI_chicken)
  e.g.: MS:1001289 (database IPI_cow)
  e.g.: MS:1001290 (database IPI_arabidopsis)
  e.g.: MS:1002060 (database UniProtKB/TrEMBL)
```

Example cvParams:

```
<cvParam cvRef="PSI-MS" accession="MS:1001013" name="database name"
value="SwissProtHuman554TD_20080526.fasta"/>
<cvParam cvRef="MS" accession="MS:1001084" name="database nr"/>
```

Example userParams:

```
<userParam name="sgd_orfs_plus_ups_prots.fasta"/>
<userParam name="itraq_db.fasta"/>
<userParam name="Pseudomonas.fasta"/>
```

6.18 Element <DataMatrix>

Definition: A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Type: DataMatrixType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Row	1	unbounded	One row of data in a data matrix.

Example Context:

```
<DataMatrix>
    <Row object_ref="pep_GAPEIDVLEGETDTK_2_21711">9079.67 6680.893 6869.919 7069.431 6872.01
7488.588 5045.095 7063.636 6636.347 3902.41 4320.357 4045.954</Row>
    <Row object_ref="pep_QSTTFADCPVVPADPDILLAK_2_48178">3299.727 576.473 43.736 1858.257 657.126
933.167 2811.395 1563.02 2621.453 3516.591 2474.139 1086.323</Row>
    <Row object_ref="pep_QKYDADVLDK_3_2006">48572.697 36224.374 49154.991 48022.429 42519.467
43865.915 33691.767 30520.391 29488.844 31172.178 23221.642 20119.87</Row>
    <Row object_ref="pep_GLGNPLLYDGVER_2_13819">14041.079 10370.307 11493.51 12698.35 9696.061
10021.047 9707.69 10754.69 10504.175 8528.331 7383.964 6172.554</Row>
    <Row object_ref="pep_LFAGMSPEMAK_2_5761">25030.664 17899.384 19336.846 26564.901 19449.275
22901.059 15937.055 21527.449 16656.611 13495.235 12320.256 12043.633</Row>
    <Row object_ref="pep_FSPVSTASYR_2_7384">15736.005 12792.928 12804.118 14132.755 14456.028
15512.023 8811.88 11793.098 10181.587 7227.684 7532.051 6258.705</Row>
  ...
</DataMatrix>
```

6.19 Element <DataProcessing>

Definition: Description of the way in which a particular software package was used to analyse data and for example produce different quant layers or lists in the file.

Type: DataProcessingType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
order	xsd:positiveInteger	required	This attributes allows a series of consecutive steps to be placed in the correct order, start counting from 1.
software_ref	xsd:IDREF	required	This attribute MUST reference the appropriate SoftwareType.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
InputObject_refs	0	1	References to raw file groups, FeatureList, PeptideConsensusList, ProteinList or ProteinGroupList or QuantLayers that were inputs to the data processing step.
OutputObject_refs	0	1	References to raw file groups, FeatureList, PeptideConsensusList, ProteinList or ProteinGroupList or QuantLayers that were the outputs of the data processing step.
ProcessingMethod	1	unbounded	Description of one step within the data processing pipeline.

Example Context:

```
<DataProcessing order="1" software_ref="x-Tracker" id="DP1">
    <ProcessingMethod order="1">
        <userParam value="load identification" name="Plugin type"/>
        <userParam value="loadMzIdentML" name="Plugin name"/>
        <userParam  value="examples/paper_iTraq4plex/loadMzIdentML.xtp"  name="Plugin  configuration file"/>
    </ProcessingMethod>
    <ProcessingMethod order="2">
  ...
</DataProcessing>
```

6.20 Element <DataProcessingList>

Definition: List and descriptions of data processing applied to this data.

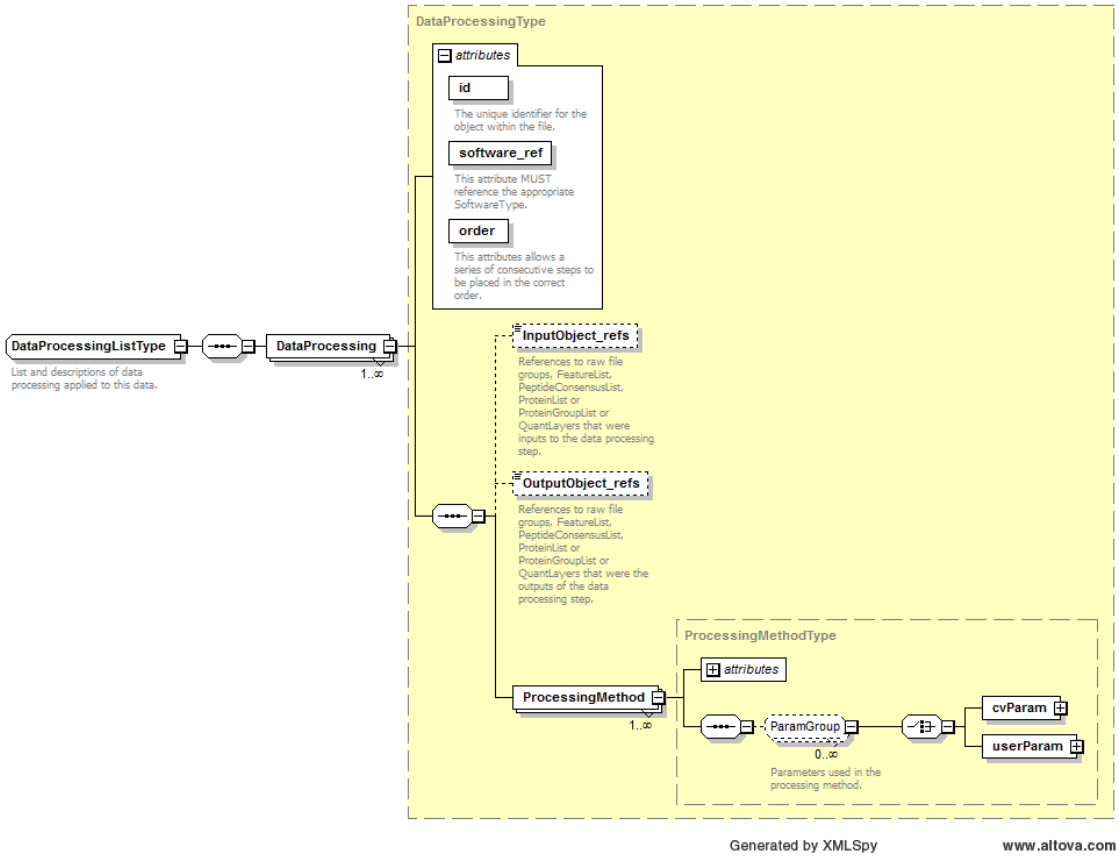
Type: DataProcessingListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataProcessing	1	unbounded	Description of the way in which a particular software package was used to analyse data and for example produce different quant layers or lists in the file.

Graphical Context:



Example Context:

```
<DataProcessingList>
    <DataProcessing order="1" software_ref="x-Tracker" id="DP1">
        <ProcessingMethod order="1">
            <userParam value="load identification" name="Plugin type"/>
            <userParam value="loadMzIdentML" name="Plugin name"/>
            <userParam  value="examples/paper_iTraq4plex/loadMzIdentML.xtp"  name="Plugin  configuration
file"/>
        </ProcessingMethod>
  ...
</DataProcessingList>
```

6.21 Element <DataType>

Definition:

Depending on context:

1: Type of data in the quant layer e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFold Change|...."

2: Type of data in the column e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance".

Type: cvParamRefType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
<DataType>
    <cvParam accession="MS:1001890" cvRef="PSI-MS" name="Progenesis:protein normalised abundance"/>
</DataType>
```

cvParam

```
Path /MzQuantML/SmallMoleculeList/GlobalQuantLayer/ColumnDefinition/Column/DataType
```

Mapping Rules:

MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinGroupList/AssayQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinGroupList/GlobalQuantLayer/ColumnDefinition/Column/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)

e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/FeatureList/MS2AssayQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/PeptideConsensusList/AssayQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinGroupList/StudyVariableQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)

e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinList/GlobalQuantLayer/ColumnDefinition/Column/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/SmallMoleculeList/StudyVariableQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/FeatureList/FeatureQuantLayer/ColumnDefinition/Column/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)

e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/FeatureList/MS2StudyVariableQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinList/StudyVariableQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/SmallMoleculeList/AssayQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)

e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/PeptideConsensusList/StudyVariableQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/ProteinList/AssayQuantLayer/DataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.
Path /MzQuantML/PeptideConsensusList/GlobalQuantLayer/ColumnDefinition/Column/DataType

MUST supply a *child* term of MS:1001805 (quantification datatype) only once
e.g.: MS:1001130 (peptide raw area)
e.g.: MS:1001131 (error on peptide area)
e.g.: MS:1001132 (peptide ratio)
e.g.: MS:1001133 (error on peptide ratio)
e.g.: MS:1001134 (protein ratio)
e.g.: MS:1001135 (error on protein ratio)
e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
e.g.: MS:1001137 (absolute quantity)
e.g.: MS:1001138 (error on absolute quantity)
e.g.: MS:1001141 (intensity of precursor ion)
et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
e.g.: MS:1000796 (spectrum title)
e.g.: MS:1000797 (peak list scans)
e.g.: MS:1000798 (peak list raw scans)
e.g.: MS:1000903 (product ion series ordinal)
e.g.: MS:1000904 (product ion m/z delta)
e.g.: MS:1000926 (product interpretation rank)
e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
e.g.: MS:1001035 (date / time search performed)
e.g.: MS:1001036 (search time taken)
e.g.: MS:1001088 (protein description)
et al.

Example cvParams:

```
<cvParam accession="MS:1001890" cvRef="PSI-MS" name="Progenesis:protein normalised abundance"/>
<cvParam accession="MS:1001892" cvRef="PSI-MS" name="Progenesis:protein raw abundance"/>
<cvParam accession="MS:1001171" cvRef="PSI-MS" name="Mascot:score"/>
<cvParam accession="MS:1001891" cvRef="PSI-MS" name="Progenesis:peptide normalised abundance"/>
<cvParam accession="MS:1001893" cvRef="PSI-MS" name="Progenesis:peptide raw abundance"/>
<cvParam accession="MS:1001903" cvRef="PSI-MS" name="MaxQuant:feature intensity"/>
<cvParam accession="MS:1001897" cvRef="PSI-MS" name="MaxQuant:peptide counts (unique)"/>
<cvParam cvRef="PSI-MS" accession="MS:1001141" name="intensity of precursor ion"/>
<cvParam cvRef="PSI-MS" accession="MS:1001105" name="peptide result details"/>
<cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
<cvParam accession="MS:1001905" cvRef="PSI-MS" name="emPAI value"/>
<cvParam cvRef="MS" accession="MS:1001860" name="protein value: mean of peptide ratios"/>
<cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
<cvParam cvRef="MS" accession="MS:1001852" name="normalized protein value"/>
```

6.22 Element <DBIdentificationRef>

Definition: External database references for the small molecule identification.

Type: DBIdentificationRefType

Attributes:

Attribute Name	Data Type	Use	Definition
id_ref	xsd:string	required	Reference to the unique identifier of this object in the referenced external file.
searchDatabase_ref	xsd:IDREF	required	Reference to the SearchDatabase object in this file from which this identification was made.

Subelements: none

Example Context:

6.23 Element <DenominatorDataType>

Definition: Type of data used for the denominator of the ratio e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFold Change|...."

Type: cvParamRefType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
<DenominatorDataType>
  <cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
</DenominatorDataType>
```

cvParam Mapping Rules:

```
Path /MzQuantML/RatioList/Ratio/DenominatorDataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
  e.g.: MS:1001130 (peptide raw area)
  e.g.: MS:1001131 (error on peptide area)
  e.g.: MS:1001132 (peptide ratio)
  e.g.: MS:1001133 (error on peptide ratio)
  e.g.: MS:1001134 (protein ratio)
  e.g.: MS:1001135 (error on protein ratio)
  e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
  e.g.: MS:1001137 (absolute quantity)
  e.g.: MS:1001138 (error on absolute quantity)
  e.g.: MS:1001141 (intensity of precursor ion)
  et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
  e.g.: MS:1000796 (spectrum title)
  e.g.: MS:1000797 (peak list scans)
  e.g.: MS:1000798 (peak list raw scans)
  e.g.: MS:1000903 (product ion series ordinal)
  e.g.: MS:1000904 (product ion m/z delta)
  e.g.: MS:1000926 (product interpretation rank)
  e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
  e.g.: MS:1001035 (date / time search performed)
  e.g.: MS:1001036 (search time taken)
  e.g.: MS:1001088 (protein description)
  et al.
```

Example cvParams:

```
<cvParam accession="MS:1001903" cvRef="PSI-MS" name="MaxQuant:feature intensity"/>
<cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
<cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
```

6.24 Element <EvidenceRef>

Definition: Evidence associated with the PeptideConsensus, including mandatory associations to features and optional references to identifications that have been assigned to the feature.

Type: EvidenceRefType

Attributes:

Attribute Name	Data Type	Use	Definition
assay_refs	xsd:IDREFS	required	A reference to the Assay to which the referenced Feature belongs (e.g. in label-free analyses) or has been assigned (e.g. in label-based analyses). Multiple values MUST only be given for MS2-tag approaches where the same feature quantifies multiple assays.
feature_ref	xsd:IDREF	required	The feature to which this identification has been assigned.
id_refs	listOfStrings	optional	One or more reference to the unique identifiers for objects identifying the relevant Peptide in an external file. This attribute MUST be present if an identificationFile_ref is given.

identificationFile_ref	xsd:IDREF	optional	Reference to the IdentificationFile object in this file. This attribute MUST be present if id_refs are given.

Subelements: none

Example Context:

```
<EvidenceRef feature_ref="ft_216" identificationFile_ref="idfile_1" id_refs="SII_69413_1"
assay_refs="ass_0"/>
```

6.25 Element <ExternalFormatDocumentation>

Definition: A URI to access documentation and tools to interpret the external format of the ExternalData instance. For example, XML Schema or static libraries (APIs) to access binary formats.

Type: xsd:anyURI

Attributes: none

Subelements: none

Example Context:

6.26 Element <Feature>

Definition: A region on a (potentially) two-dimensional map of LC-MS (MS1) scans, defined by the retention time, mass over charge and optionally a mass trace. Quantitative values about features can be added in the associated QuantLayers. For techniques that analyse data from single scans e.g. MS2 tagging approaches, a Feature corresponds with the mz of the parent ions only.

Type: FeatureType

Attributes:

Attribute Name	Data Type	Use	Definition
charge	integerOrNullType	required	The assumed charge of the feature, used to calculate the m/z value. If the charge is unknown the null type MUST be used.
chromatogram_refs	xsd:string	optional	The identifier(s) of the chromatogram from which this feature was detected, for example required for SRM analyses. Multiple values can be provided separated by spaces following the XSD:IDREFS style.
id	xsd:ID	required	Unique identifier for the feature.
mz	xsd:double	required	The monoisotopic peak on the mass over charge axis in daltons over charge.
rawFile_ref	xsd:IDREF	optional	An optional reference to a RawFile element inside a RawFileGroup. The reference MUST be used if multiple RawFiles are provided within a RawFilesGroup, e.g. if pre-fractionation has occurred, and SHOULD not be used otherwise.
rt	doubleOrNullType	required	The centre point of the feature on the retention time axis in minutes. If the retention time is unknown, the value should be null.

spectrum_refs	xsd:string	optional	An optional identifier for an individual spectrum from which this feature was detected, required only for techniques where rt mapping is not appropriate. Multiple values can be provided separated by spaces following the XSD:IDREFS style.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
MassTrace	0	1	The coordinates defining the feature in RT and MZ space, given as boundary points or a series of rectangles, as encoded by the MassTraceEncoding cvParam on the FeatureList.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<Feature id="f_14215451996852426123" rt="5002.86467743247" mz="580.976424828825" charge="3">
  <userParam name="map_index" value="0"/>
  <userParam name="feature_index" value="0"/>
</Feature>
```

cvParam Mapping Rules:

```
Path /MzQuantML/FeatureList/Feature
MAY supply a *child* term of MS:1001828 (feature attribute) one or more times
  e.g.: MS:1001829 (SRM transition ID)
  e.g.: MS:1002121 (spectral count feature)
```

Example userParams:

```
<userParam name="map_index" value="0"/>
<userParam name="feature_index" value="0"/>
```

6.27 Element <FeatureList>

Definition: All the features identified on a single raw file or raw file group.

Type: FeatureListType

Attributes:

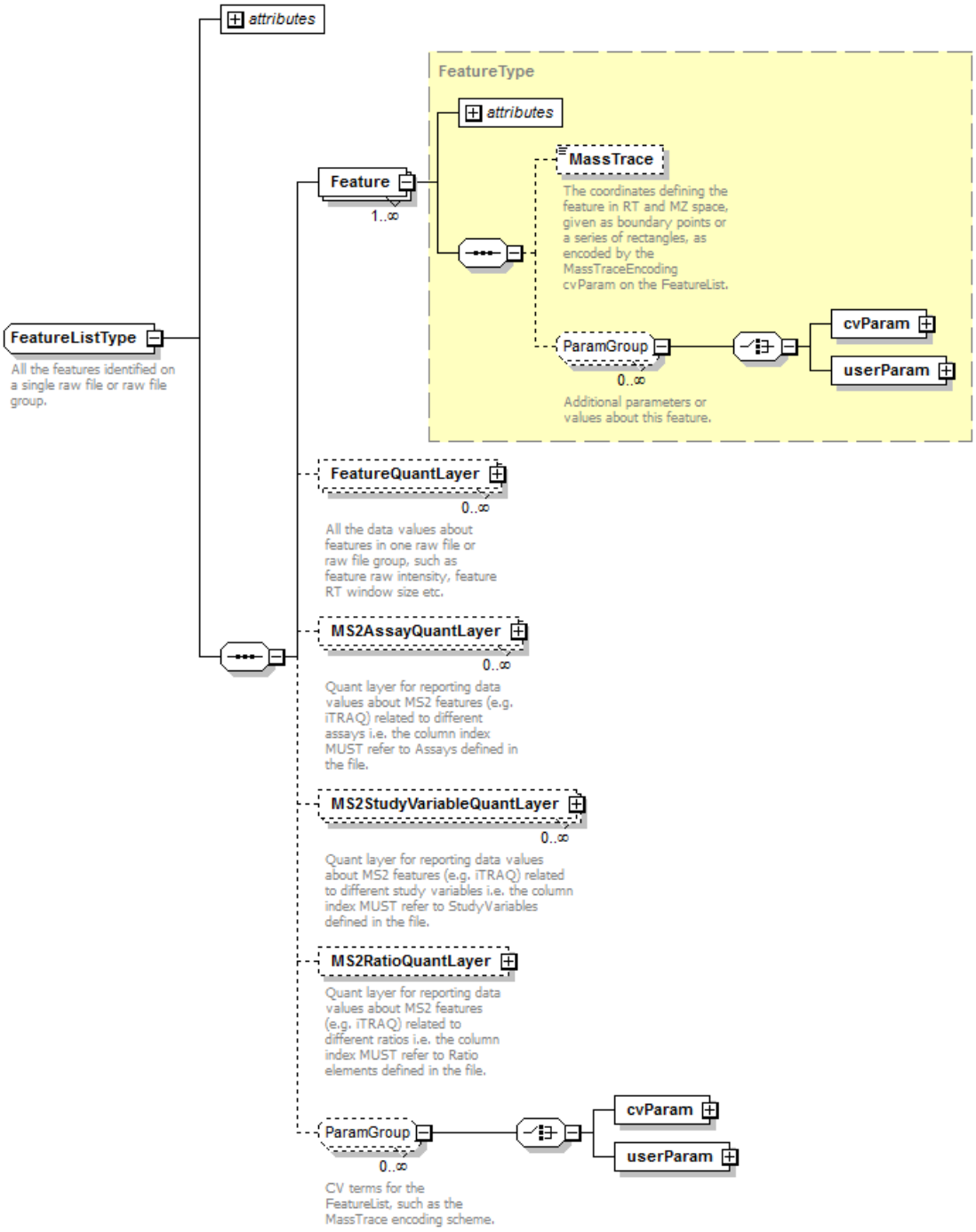
Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
rawFilesGroup_ref	xsd:IDREF	required	Reference to the raw file or group of raw files from which this feature list was generated.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Feature	1	unbounded	A region on a (potentially) two-dimensional map of LC-MS (MS1) scans, defined by the retention time, mass over charge and optionally a mass trace. Quantitative values about features can be added in the associated QuantLayers. For techniques that analyse data from single scans e.g. MS2 tagging approaches, a Feature corresponds with the mz of the parent ions only.

FeatureQuantLayer	0	unbounded	All the data values about features in one raw file or raw file group, such as feature raw intensity, feature RT window size etc.
MS2AssayQuantLayer	0	unbounded	Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different assays i.e. the column index MUST refer to Assays defined in the file.
MS2StudyVariableQuantLayer	0	unbounded	Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.
MS2RatioQuantLayer	0	1	Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary (e.g. the encoding of the MassTrace).
userParam	0	unbounded	A single user-defined parameter.

Graphical Context:



Example Context:

```
<FeatureList id="featureList_raw1" rawFilesGroup_ref="raw1">
    <Feature spectrum_refs="Scan Number: 1656" charge="2" mz="411.6925" rt="null" id="raw1-PA0958-
SADFIGGR_00000000-2-411.6925_Scan_Number_1656"/>
    <Feature spectrum_refs="Scan Number: 1205" charge="2" mz="474.7568" rt="null" id="raw1-PA0958-
YVVQSGPAK_000000000-2-474.7568_Scan_Number_1205"/>
    <Feature spectrum_refs="Scan Number: 3667" charge="2" mz="573.2433" rt="null" id="raw1-PA0958-
LIVDYPLSIL_0000000000-2-573.2433_Scan_Number_3667"/>
    <Feature spectrum_refs="Scan Number: 3612" charge="2" mz="573.2534" rt="null" id="raw1-PA0958-
LIVDYPLSIL_0000000000-2-573.2534_Scan_Number_3612"/>
    <Feature spectrum_refs="Scan Number: 1995" charge="2" mz="687.29" rt="null" id="raw1-PA0958-
GELYATYAGETAK_0000000000000-2-687.29_Scan_Number_1995"/>
    <Feature spectrum_refs="Scan Number: 1877" charge="2" mz="687.3034" rt="null" id="raw1-PA0958-
GELYATYAGETAK_0000000000000-2-687.3034_Scan_Number_1877"/>
  ...
```

```
</FeatureList>
```

cvParam Mapping Rules:

```
Path /MzQuantML/FeatureList
MAY supply a *child* term of MS:1001825 (feature list attribute) one or more times
  e.g.: MS:1001826 (mass trace reporting: rectangles)
  e.g.: MS:1001827 (mass trace reporting: polygons)
  e.g.: MS:1002122 (counts reporting)
```

6.28 Element <FeatureQuantLayer>

Definition: All the data values about features in one raw file or raw file group, such as feature raw intensity, feature RT window size etc.

Type: GlobalQuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ColumnDefinition	1	1	Definition of the data types in each column.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

```
    <FeatureQuantLayer id="q_14611336223554821348">
      <ColumnDefinition>
        <Column index="0">
          <DataType>
            <cvParam cvRef="PSI-MS" accession="MS:1001141" name="intensity of precursor ion"/>
          </DataType>
        </Column>
        <Column index="1">
  ...
</FeatureQuantLayer>
```

6.29 Element <Feature_refs>

Definition: Optional references to features on which quantification values about the SmallMolecule in the QuantLayer were based.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

6.30 Element <FileFormat>

Definition: The format of the ExternalData file, for example "tiff" for image files.

Type: FileFormatType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
<FileFormat>
    <cvParam cvRef="MS" accession="MS:1000914" name="tab delimited text file"/>
</FileFormat>
```

cvParam Mapping Rules:

Path /MzQuantML/InputFiles/RawFilesGroup/RawFile/FileFormat
MUST supply a *child* term of MS:1000560 (mass spectrometer file format) one or more times
  e.g.: MS:1000526 (Waters raw file)
  e.g.: MS:1000562 (ABI WIFF file)
  e.g.: MS:1000563 (Thermo RAW file)
  e.g.: MS:1000564 (PSI mzData file)
  e.g.: MS:1000565 (Micromass PKL file)
  e.g.: MS:1000566 (ISB mzXML file)
  e.g.: MS:1000567 (Bruker/Agilent YEP file)
  e.g.: MS:1000584 (mzML file)
  e.g.: MS:1000613 (DTA file)
  e.g.: MS:1000614 (ProteinLynx Global Server mass spectrum XML file)
  et al.
Path /MzQuantML/InputFiles/MethodFiles/MethodFile/FileFormat
MAY supply a *child* term of MS:1002128 (method file format attribute) only once
Path /MzQuantML/InputFiles/IdentificationFiles/IdentificationFile/FileFormat
MUST supply a *child* term of MS:1001040 (intermediate analysis format) only once
  e.g.: MS:1000742 (Bioworks SRF file)
  e.g.: MS:1000914 (tab delimited text file)
  e.g.: MS:1001107 (data stored in database)
  e.g.: MS:1001199 (Mascot DAT file)
  e.g.: MS:1001200 (Sequest out file)
  e.g.: MS:1001242 (Sequest out folder)
  e.g.: MS:1001243 (Sequest summary)
  e.g.: MS:1001275 (ProteinScape SearchEvent)
  e.g.: MS:1001276 (ProteinScape Gel)
  e.g.: MS:1001399 (OMSSA csv file)
  et al.
Path /MzQuantML/InputFiles/SourceFile/FileFormat
MUST supply a *child* term of MS:1001040 (intermediate analysis format) only once
  e.g.: MS:1000742 (Bioworks SRF file)
  e.g.: MS:1000914 (tab delimited text file)
  e.g.: MS:1001107 (data stored in database)
  e.g.: MS:1001199 (Mascot DAT file)
  e.g.: MS:1001200 (Sequest out file)
  e.g.: MS:1001242 (Sequest out folder)
  e.g.: MS:1001243 (Sequest summary)
  e.g.: MS:1001275 (ProteinScape SearchEvent)
  e.g.: MS:1001276 (ProteinScape Gel)
  e.g.: MS:1001399 (OMSSA csv file)
  et al.
Path /MzQuantML/InputFiles/SearchDatabase/FileFormat
MUST supply a *child* term of MS:1001347 (database file formats) only once
  e.g.: MS:1001348 (FASTA format)
  e.g.: MS:1001349 (ASN.1)
  e.g.: MS:1001350 (NCBI *.p*)
  e.g.: MS:1001351 (clustal aln)
  e.g.: MS:1001352 (embl em)
  e.g.: MS:1001353 (NBRF PIR)
  e.g.: MS:1001462 (PEFF format)

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000914" name="tab delimited text file"/>
```

6.31 Element <GlobalQuantLayer>

Definition:

Depending on context:

1: Global values corresponding to the ProteinGroup such as the total intensity of the protein group in all assays, Anova etc.

2: Global values corresponding to the Protein such as the total intensity of the protein in all assays, Anova etc.

3: Global values corresponding to the Peptide such as the total intensity of peptide in all assays, Anova in a quantitative peptidome experiment etc.

4: Global values corresponding to the small molecule such as the total intensity of the molecule

in all assays, Anova etc.

Type: GlobalQuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ColumnDefinition	1	1	Definition of the data types in each column.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

```
<GlobalQuantLayer id="Pep_GQL1">
    <ColumnDefinition>
        <Column index="0">
            <DataType>
                <cvParam accession="MS:1001171" cvRef="PSI-MS" name="Mascot:score"/>
            </DataType>
        </Column>
  ...
</GlobalQuantLayer>
```

6.32 Element <IdentificationFile>

Definition: A single identification file associated with this analysis.

Type: IdentificationFileType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
location	xsd:anyURI	required	The location of the data file.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.
searchDatabase_ref	xsd:IDREF	optional	Reference to the SearchDatabase object in this file from which this identification was made.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ExternalFormatDocumentation	0	1	A URI to access documentation and tools to interpret the external format of the ExternalData instance. For example, XML Schema or static libraries (APIs) to access binary formats.
FileFormat	0	1	The format of the ExternalData file, for example "tiff" for image files.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<IdentificationFile location="examples/paper_iTraq4plex/itraq.mzid" id="id_file1"/>
```

cvParam Mapping Rules:

```
Path /MzQuantML/InputFiles/IdentificationFiles/IdentificationFile
MAY supply a *child* term of MS:1002127 (identification file attribute) one or more times
```

6.33 Element <IdentificationFiles>

Definition: All identification files associated with this quantitation analysis to be referenced elsewhere.

Type: IdentificationFilesType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
IdentificationFile	1	unbounded	A single identification file associated with this analysis.

Example Context:

```
<IdentificationFiles>
            <IdentificationFile location="examples/paper_iTraq4plex/itraq.mzid" id="id_file1"/>
</IdentificationFiles>
```

6.34 Element <IdentificationFile_refs>

Definition: One or more identification files used within this assay.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

6.35 Element <IdentificationRef>

Definition:
Depending on context:
1: Reference for the identification evidence for peptides from the referenced external file and unique identifier e.g. a link to an mzIdentML file and ID for the ProteinAmbiguityGroup.
2: Reference for the identification evidence for peptides from the referenced external file and unique identifier e.g. a link to an mzIdentML file and ID for the ProteinDetectionHypothesis.

Type: IdentificationRefType

Attributes:

Attribute Name	Data Type	Use	Definition
id_refs	listOfStrings	required	One or more reference to the unique identifiers for objects identifying the relevant Peptide, Protein or Protein group in an external file.
identificationFile_ref	xsd:IDREF	required	Reference to the IdentificationFile object in this file.

Subelements: none

Example Context:

6.36 Element <InputFiles>

Definition: All the raw files, identification files and databases used in the quantitation.

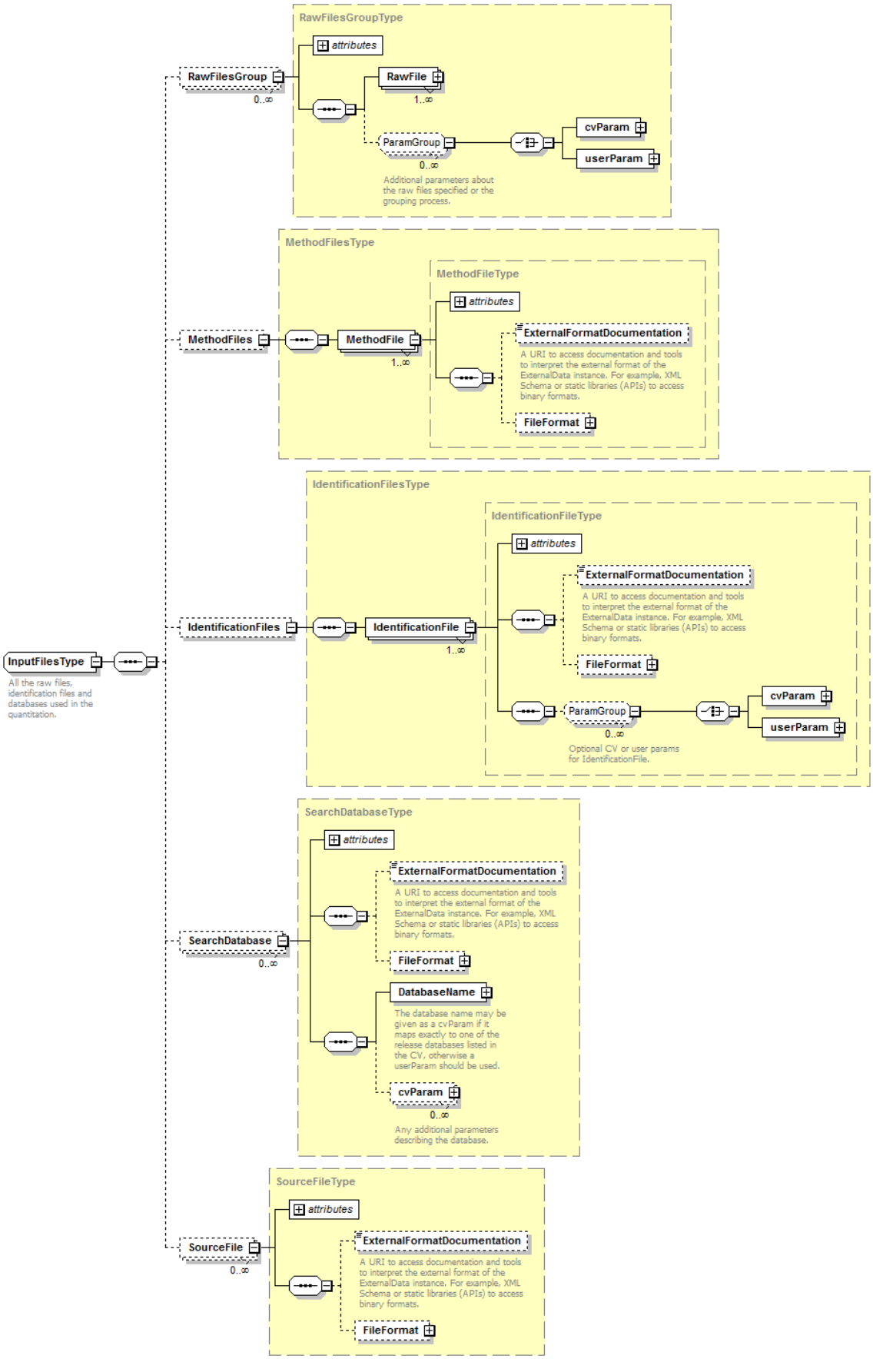
Type: InputFilesType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
RawFilesGroup	0	unbounded	The raw file or collection of raw files that together form one unit of analysis. This is mandatory unless raw files were not used for quantitation e.g. spectral counting. Multiple raw files should only be provided within a group if they have been used for sample pre-fractionation which are later summed together.
MethodFiles	0	1	All methods files associated with this quantitation analysis to be referenced elsewhere, for example used to describe TraML files used in SRM analysis.
IdentificationFiles	0	1	All identification files associated with this quantitation analysis to be referenced elsewhere.
SearchDatabase	0	unbounded	A database used for searching mass spectra. Examples include a set of amino acid sequence entries, or annotated spectra libraries.
SourceFile	0	unbounded	A file from which this MzQuantML instance was created, including potentially MzQuantML files for earlier stages in a workflow.

Graphical Context:



Example Context:

```
<InputFiles>
    <RawFilesGroup id="rg_0">
        <RawFile location="../msmsdata/mam_042408o_CPTAC_study6_6B011.raw"
            name="mam_042408o_CPTAC_study6_6B011.raw" id="raw_0"/>
    </RawFilesGroup>
    <RawFilesGroup id="rg_1">
        <RawFile location="../msmsdata/mam_050108o_CPTAC_study6_6B011.raw"
            name="mam_050108o_CPTAC_study6_6B011.raw" id="raw_1"/>
    </RawFilesGroup>
  ...
</InputFiles>
```

6.37 Element <InputObject_refs>

Definition: References to raw file groups, FeatureList, PeptideConsensusList, ProteinList or ProteinGroupList or QuantLayers that were inputs to the data processing step.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

```
<InputObject_refs>AQL_MEANS_OF_GROUPS</InputObject_refs>
```

6.38 Element <Label>

Definition: A specification of labels or tags used to define the assay within the raw file, such as heavy labelling or iTRAQ tag mass. The Label and Modification is mandatory so a specific term is provided under Modification for unlabeled sample for label-free and, for example, so-called light samples in a labelling experiment.

Type: LabelType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Modification	1	unbounded	The modification searched for or used to define the label or tag for quantification. It SHOULD be sourced from UniMod, PSI-MS or PSI-MOD.

Example Context:

```
      <Label>
        <Modification massDelta="8.0141988132" >
          <cvParam cvRef="PSI-MOD" accession="MOD:00582" name="6x(13)C,2x(15)N labeled L-lysine"
value="Lys8"/>
        </Modification>
        <Modification massDelta="10.0082686" >
          <cvParam cvRef="PSI-MOD" accession="MOD:00587" name="6x(13)C,4x(15)N labeled L-arginine"
value="Arg10"/>
        </Modification>
  ...
</Label>
```

6.39 Element <MassTrace>

Definition: The coordinates defining the feature in RT and MZ space, given as boundary points or a series of rectangles, as encoded by the MassTraceEncoding cvParam on the FeatureList.

Type: psi-mzq:listOfDoubles

Attributes: none

Subelements: none

Example Context:

```
<MassTrace>76.344 1129.574 76.916 1131.074</MassTrace>
```

Example cvParams:

```
<cvParam accession="MS:1001826" cvRef="PSI-MS" name="mass trace reporting: rectangles"/>
```

6.40 Element <MethodFile>

Definition: A single methods file associated with this analysis e.g. a TraML file used for SRM analysis.

Type: MethodFileType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
location	xsd:anyURI	required	The location of the data file.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ExternalFormatDocumentation	0	1	A URI to access documentation and tools to interpret the external format of the ExternalData instance. For example, XML Schema or static libraries (APIs) to access binary formats.
FileFormat	0	1	The format of the ExternalData file, for example "tiff" for image files.

Example Context:

6.41 Element <MethodFiles>

Definition: All methods files associated with this quantitation analysis to be referenced elsewhere, for example used to describe TraML files used in SRM analysis.

Type: MethodFilesType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
MethodFile	1	unbounded	A single methods file associated with this analysis e.g. a TraML file used for SRM analysis.

Example Context:

6.42 Element <Modification>

Definition: The modification searched for or used to define the label or tag for quantification. It SHOULD be sourced from UniMod, PSI-MS or PSI-MOD.

Type: ModParamType

Attributes:

Attribute Name	Data Type	Use	Definition
massDelta	xsd:float	optional	The mass delta of the modification in Daltons
residues	listOfChars	optional	The residue(s) to which the modification has been applied.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

6.42.1 Specification of a <Modification> searched for

Example Context:

```
<Modification monoisotopicMassDelta="15.994915" location="13">
    <cvParam accession="UNIMOD:35" cvRef="UNIMOD" name="Oxidation"/>
    <cvParam accession="MS:1001524" cvRef="PSI-MS" value="63.998285" name="fragment neutral loss"/>
</Modification>
```

cvParam Mapping Rules:

```
Path /MzQuantML/PeptideConsensusList/PeptideConsensus/Modification
MAY supply a *child* term of UNIMOD:0 (unimod root node) only once
MAY supply a *child* term of MS:1001471 (peptide modification details) only once
  e.g.: MS:1001460 (unknown modification)
  e.g.: MS:1001524 (fragment neutral loss)
  e.g.: MS:1001525 (precursor neutral loss)
  e.g.: MS:1001972 (PTM scoring algorithm version)
  e.g.: MS:1002028 (nucleic acid base modification)
  e.g.: MS:1002029 (original nucleic acid sequence)
  e.g.: MS:1002030 (modified nucleic acid sequence)
MAY supply a *child* term of MOD:00000 (protein modification) only once
```

Example cvParams:

```
<cvParam accession="UNIMOD:35" cvRef="UNIMOD" name="Oxidation"/>
<cvParam accession="MS:1001524" cvRef="PSI-MS" value="63.998285" name="fragment neutral loss"/>
```

6.42.2 Modifications used to define a label or tag for quantification

6.42.2.1 For multiplex techniques, such as a 4plex iTRAQ analysis

Example Context:

```
<Modification massDelta="145.0">
     <cvParam accession="MOD:01522" cvRef="PSI-MOD" value="114" name="iTRAQ4plex-114 reporter fragment"/>
</Modification>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AssayList/Assay/Label/Modification
MUST supply a *child* term of UNIMOD:0 (unimod root node) only once
MUST supply a *child* term of MS:1001471 (peptide modification details) only once
  e.g.: MS:1001460 (unknown modification)
  e.g.: MS:1001524 (fragment neutral loss)
  e.g.: MS:1001525 (precursor neutral loss)
  e.g.: MS:1001972 (PTM scoring algorithm version)
  e.g.: MS:1002028 (nucleic acid base modification)
  e.g.: MS:1002029 (original nucleic acid sequence)
  e.g.: MS:1002030 (modified nucleic acid sequence)
MUST supply term MS:1002038 (unlabeled sample) only once
MUST supply a *child* term of MOD:00000 (protein modification) only once
```

Example cvParams:

```
<cvParam accession="MOD:01522" cvRef="PSI-MOD" value="114" name="iTRAQ4plex-114 reporter fragment"/>
<cvParam accession="MOD:01523" cvRef="PSI-MOD" value="115" name="iTRAQ4plex-115 reporter
```

```
fragment"/>
<cvParam accession="MOD:01524" cvRef="PSI-MOD" value="116" name="iTRAQ4plex-116 reporter
fragment"/>
<cvParam accession="MOD:01525" cvRef="PSI-MOD" value="117" name="iTRAQ4plex-117, mTRAQ heavy,
reporter fragment"/>
<cvParam accession="UNIMOD:214" cvRef="UNIMOD" name="iTRAQ4plex"/>
```

6.42.2.2 For a 2plex SILAC, experiment

Example Context:

```
<Modification massDelta="8.0141988132" >
        <cvParam  cvRef="PSI-MOD"  accession="MOD:00582"  name="6x(13)C,2x(15)N  labeled  L-
lysine" value="Lys8"/>
</Modification>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AssayList/Assay/Label/Modification
MUST supply a *child* term of UNIMOD:0 (unimod root node) only once
MUST supply a *child* term of MS:1001471 (peptide modification details) only once
  e.g.: MS:1001460 (unknown modification)
  e.g.: MS:1001524 (fragment neutral loss)
  e.g.: MS:1001525 (precursor neutral loss)
  e.g.: MS:1001972 (PTM scoring algorithm version)
  e.g.: MS:1002028 (nucleic acid base modification)
  e.g.: MS:1002029 (original nucleic acid sequence)
  e.g.: MS:1002030 (modified nucleic acid sequence)
MUST supply term MS:1002038 (unlabeled sample) only once
MUST supply a *child* term of MOD:00000 (protein modification) only once
```

Example cvParams:

```
<cvParam accession="MS:1002038" cvRef="PSI-MS" name="unlabeled sample"/>
<cvParam accession="MOD:00582" cvRef="PSI-MOD" value="Lys8" name="6x(13)C,2x(15)N labeled L-
lysine"/>
<cvParam accession="MOD:00587" cvRef="PSI-MOD" value="Arg10" name="6x(13)C,4x(15)N labeled L-
arginine"/>
```

6.42.2.3 In label-free analyses

Example Context:

```
<Modification massDelta="0" >
      <cvParam cvRef="PSI-MOD" accession="MS:1002038" name="unlabeled sample" value="none"/>
</Modification>
```

Example cvParams:

```
<cvParam accession="MS:1002038" cvRef="PSI-MS" name="unlabeled sample"/>
```

6.42.3 Modifications in Small Molecules

cvParam Mapping Rules:

```
Path /MzQuantML/SmallMoleculeList/SmallMolecule/Modification
MAY supply a *child* term of MS:1002119 (small molecule modification attribute) only once
```

6.43 Element <MS2AssayQuantLayer>

Definition: Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different assays i.e. the column index MUST refer to Assays defined in the file.

Type: QuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataType	1	1	Type of data in the column e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance".
ColumnIndex	1	1	Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.
DataMatrix	1	1	A matrix of data stored in rows and columns, as

			defined in the parent QuantLayer.

Example Context:

```
<MS2AssayQuantLayer id="MS2AssayQuantLayer_raw1_reporter_ion_intensity">
    <DataType>
        <cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
    </DataType>
    <ColumnIndex>_114 _115 _116 _117</ColumnIndex>
    <DataMatrix>
        <Row object_ref="raw1-sp_P00924_ENO1_YEAST-SVYDSR_00000000-2-
435.7178_SIR_62">3.118874006098781 2.628302416711611 1.9418652317472234 1.8073460629379474</Row>
  ...
</MS2AssayQuantLayer>
```

6.44 Element <MS2RatioQuantLayer>

Definition: Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.

Type: RatioQuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ColumnIndex	1	1	Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

6.45 Element <MS2StudyVariableQuantLayer>

Definition: Quant layer for reporting data values about MS2 features (e.g. iTRAQ) related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

Type: QuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataType	1	1	Type of data in the column e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance".
ColumnIndex	1	1	Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example

Context:

6.46 Element <NumeratorDataType>

Definition: Type of data used for the numerator of the ratio e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFold Change|...."

Type: cvParamRefType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
<NumeratorDataType>
  <cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
</NumeratorDataType>
```

cvParam Mapping Rules:

```
Path /MzQuantML/RatioList/Ratio/NumeratorDataType
MUST supply a *child* term of MS:1001805 (quantification datatype) only once
  e.g.: MS:1001130 (peptide raw area)
  e.g.: MS:1001131 (error on peptide area)
  e.g.: MS:1001132 (peptide ratio)
  e.g.: MS:1001133 (error on peptide ratio)
  e.g.: MS:1001134 (protein ratio)
  e.g.: MS:1001135 (error on protein ratio)
  e.g.: MS:1001136 (p-value (protein diff from 1 randomly))
  e.g.: MS:1001137 (absolute quantity)
  e.g.: MS:1001138 (error on absolute quantity)
  e.g.: MS:1001141 (intensity of precursor ion)
  et al.
OR MUST supply a *child* term of MS:1001405 (spectrum identification result details) only once
  e.g.: MS:1000796 (spectrum title)
  e.g.: MS:1000797 (peak list scans)
  e.g.: MS:1000798 (peak list raw scans)
  e.g.: MS:1000903 (product ion series ordinal)
  e.g.: MS:1000904 (product ion m/z delta)
  e.g.: MS:1000926 (product interpretation rank)
  e.g.: MS:1001030 (number of peptide seqs compared to each spectrum)
  e.g.: MS:1001035 (date / time search performed)
  e.g.: MS:1001036 (search time taken)
  e.g.: MS:1001088 (protein description)
  et al.
```

Example cvParams:

```
<cvParam accession="MS:1001903" cvRef="PSI-MS" name="MaxQuant:feature intensity"/>
<cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
<cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
```

6.47 Element <Organization>

Definition: Organizations are entities like companies, universities, government agencies. Any additional information such as the address, email etc. should be supplied either as CV parameters or as user parameters.

Type: OrganizationType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.

name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.
ParentOrganization	0	1	The containing organization (the university or business which a lab belongs to, etc.)

Example Context:

```
<Organization id="ORG_MPC" name="MPC - Medizinisches Proteom Center, Bochum">
    <ParentOrganization organization_ref="ORG_Univ_Bochum"/>
</Organization>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AuditCollection/Organization
SHOULD supply term MS:1000588 (contact URL) one or more times
SHOULD supply term MS:1000587 (contact address) one or more times
SHOULD supply term MS:1000589 (contact email) one or more times
SHOULD supply term MS:1000586 (contact name) one or more times
```

6.48 Element <OutputObject_refs>

Definition: References to raw file groups, FeatureList, PeptideConsensusList, ProteinList or ProteinGroupList or QuantLayers that were the outputs of the data processing step.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

6.49 Element <ParentOrganization>

Definition: The containing organization (the university or business which a lab belongs to, etc.)

Type: ParentOrganizationType

Attributes:

Attribute Name	Data Type	Use	Definition
organization_ref	xsd:IDREF	required	A reference to the organization this contact belongs to.

Subelements: none

Example Context:

```
<ParentOrganization organization_ref="ORG_Univ_Bochum"/>
```

6.50 Element <PeptideConsensus>

Definition: An element representing a peptide in different assays that may or may not have been identified. If it has been identified, the sequence and modification(s) SHOULD be reported. Within the parent list, it is allowed for there to be multiple instances of the same peptide sequence, for example capturing different charge states or different modifications, if they are differentially quantified. If peptides with different charge states are aggregated, they should be represented by a single PeptideConsensus element.

Type: PeptideConsensusType

Attributes:

Attribute Name	Data Type	Use	Definition
charge	listOfIntegers	required	The charge of this instance of Peptide. If more than one value is provided, it is assumed that this Peptide element is summarising multiple charge states of the Peptide.
id	xsd:ID	required	The unique identifier for the object within the file.
searchDatabase_ref	xsd:IDREF	optional	The search database from which this peptide was identified

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
PeptideSequence	0	1	The amino acid sequence of the (poly)peptide. If a substitution modification has been found, the original sequence should be reported. The peptide sequence is mandatory unless this is a PeptideConsensus that has not been assigned to a peptide sequence.
Modification	0	unbounded	A molecule modification specification. If n modifications have been found on a peptide, there should be n instances of Modification. If multiple modifications are provided as cvParams, it is assumed that the modification is ambiguous i.e. one modification or another. A cvParam MUST be provided with the identification of the modification sourced from a suitable CV e.g. UNIMOD. If the modification is not present in the CV (and this will be checked by the semantic validator within a given tolerance window), there is an _unknown modification_CV term that MUST be used instead. A neutral loss should be defined as an additional CVParam within Modification.
EvidenceRef	1	unbounded	Evidence associated with the PeptideConsensus, including mandatory associations to features and optional references to identifications that have been assigned to the feature.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<PeptideConsensus searchDatabase_ref="SD1" charge="2" id="pep_GAPEIDVLEGETDTK_2_21711">
    <PeptideSequence>GAPEIDVLEGETDTK</PeptideSequence>
    <EvidenceRef feature_ref="ft_216" identificationFile_ref="idfile_1" id_refs="SII_69413_1"
        assay_refs="ass_0"/>
    <EvidenceRef feature_ref="ft_217" identificationFile_ref="idfile_1" id_refs="SII_69415_1"
        assay_refs="ass_1"/>
    <EvidenceRef feature_ref="ft_218" identificationFile_ref="idfile_1" id_refs="SII_69417_1"
        assay_refs="ass_2"/>
```

```
    <EvidenceRef feature_ref="ft_219" identificationFile_ref="idfile_1" id_refs="SII_69414_1"
        assay_refs="ass_3"/>
    <EvidenceRef feature_ref="ft_220" identificationFile_ref="idfile_1" id_refs="SII_69418_1"
        assay_refs="ass_4"/>
    <EvidenceRef feature_ref="ft_221" identificationFile_ref="idfile_1" id_refs="SII_69416_1"
        assay_refs="ass_5"/>
    <EvidenceRef feature_ref="ft_222" assay_refs="ass_6"/>
    <EvidenceRef feature_ref="ft_223" assay_refs="ass_7"/>
    <EvidenceRef feature_ref="ft_224" assay_refs="ass_8"/>
    <EvidenceRef feature_ref="ft_225" assay_refs="ass_9"/>
    <EvidenceRef feature_ref="ft_226" assay_refs="ass_10"/>
    <EvidenceRef feature_ref="ft_227" assay_refs="ass_11"/>
</PeptideConsensus>
```

cvParam Mapping Rules:

```
Path /MzQuantML/PeptideConsensusList/PeptideConsensus
MAY supply a *child* term of MS:1002116 (peptide consensus attribute) only once
```

6.51 Element <PeptideConsensusList>

Definition: The list of all peptides for which quantitation values are reported.

Type: PeptideConsensusListType

Attributes:

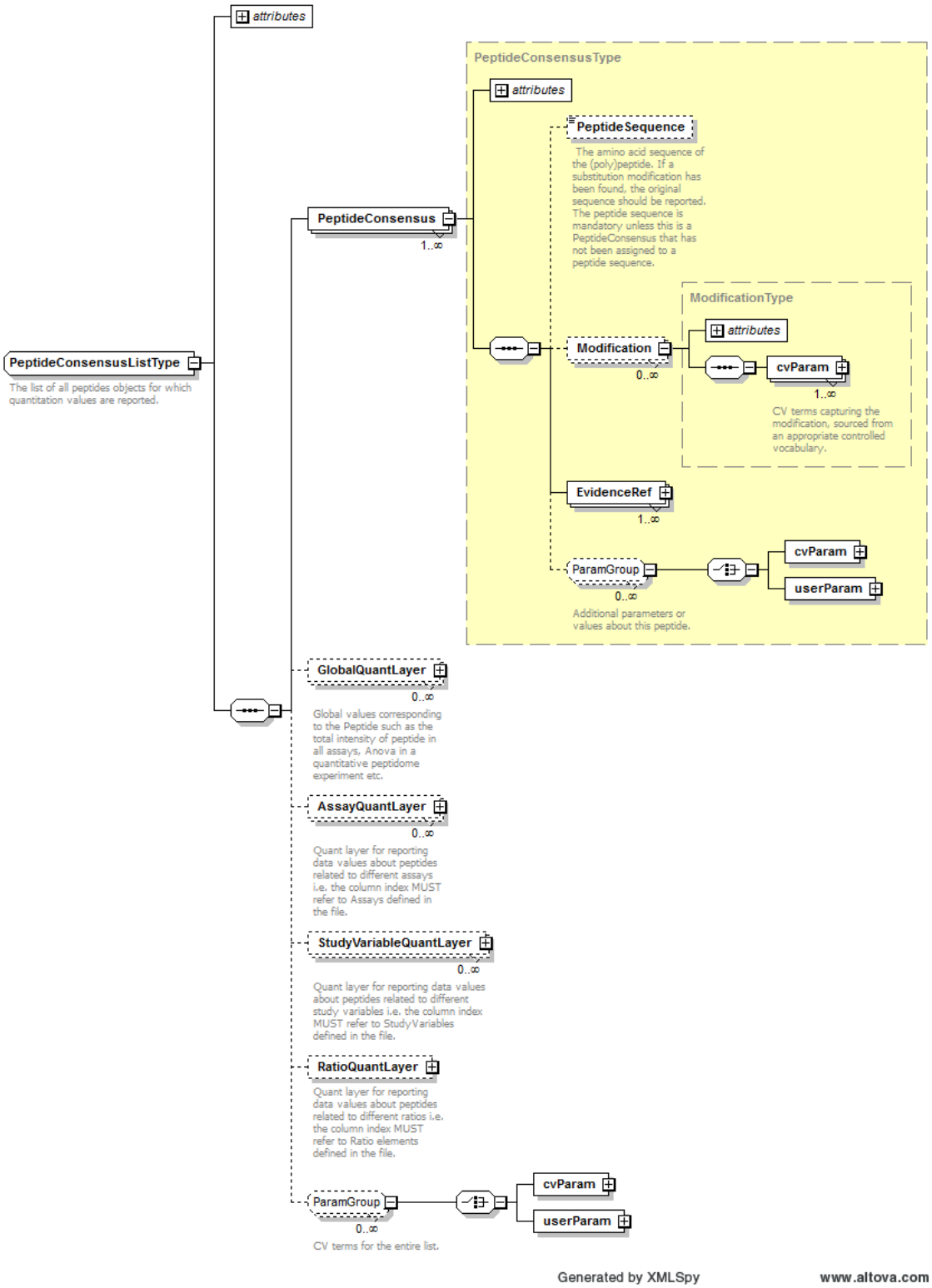
Attribute Name	Data Type	Use	Definition
finalResult	xsd:boolean	required	Multiple peptide lists are allowed for reporting the evidence trail to create a final peptide list, but the final result e.g. to be loaded into a database, MUST be flagged with finalResult=true
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
PeptideConsensus	1	unbounded	An element representing a peptide in different assay that may or may not have been identified. If it has been identified, the sequence and modification(s) SHOULD be reported. Within the parent list, it is allowed for there to be multiple instances of the same peptide sequence, for example capturing different charge states or different modifications, if they are differentially quantified. If peptides with different charge states are aggregated, they should be represented by a single PeptideConsensus element.
GlobalQuantLayer	0	unbounded	Global values corresponding to the Peptide such as the total intensity of peptide in all assays, Anova in a quantitative peptidome experiment etc.
AssayQuantLayer	0	unbounded	A collection of data relating to the objects within the parent list type (e.g. PeptideConsensus, Protein or ProteinGroup).

StudyVariableQuantLayer	0	unbounded	Quant layer for reporting data values about peptides related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.
RatioQuantLayer	0	1	Quant layer for reporting data values about peptides related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Graphical Context:



Example Context:

```
<PeptideConsensusList finalResult="true" id="PepList1">
    <PeptideConsensus searchDatabase_ref="SD1" charge="2" id="pep_GAPEIDVLEGETDTK_2_21711">
        <PeptideSequence>GAPEIDVLEGETDTK</PeptideSequence>
        <EvidenceRef feature_ref="ft_216" identificationFile_ref="idfile_1" id_refs="SII_69413_1"
            assay_refs="ass_0"/>
        <EvidenceRef feature_ref="ft_217" identificationFile_ref="idfile_1" id_refs="SII_69415_1"
            assay_refs="ass_1"/>
        <EvidenceRef feature_ref="ft_218" identificationFile_ref="idfile_1" id_refs="SII_69417_1"
            assay_refs="ass_2"/>
        <EvidenceRef feature_ref="ft_219" identificationFile_ref="idfile_1" id_refs="SII_69414_1"
            assay_refs="ass_3"/>
```

```
            <EvidenceRef feature_ref="ft_220" identificationFile_ref="idfile_1" id_refs="SII_69418_1"
                assay_refs="ass_4"/>
            <EvidenceRef feature_ref="ft_221" identificationFile_ref="idfile_1" id_refs="SII_69416_1"
                assay_refs="ass_5"/>
            <EvidenceRef feature_ref="ft_222" assay_refs="ass_6"/>
            <EvidenceRef feature_ref="ft_223" assay_refs="ass_7"/>
            <EvidenceRef feature_ref="ft_224" assay_refs="ass_8"/>
            <EvidenceRef feature_ref="ft_225" assay_refs="ass_9"/>
            <EvidenceRef feature_ref="ft_226" assay_refs="ass_10"/>
            <EvidenceRef feature_ref="ft_227" assay_refs="ass_11"/>
        </PeptideConsensus>
        <PeptideConsensus searchDatabase_ref="SD1" charge="2" id="pep_QSTTFADCPVVPADPDILLAK_2_48178">
            <PeptideSequence>QSTTFADCPVVPADPDILLAK</PeptideSequence>
            <EvidenceRef feature_ref="ft_72" identificationFile_ref="idfile_1" id_refs="SII_92174_1"
                assay_refs="ass_0"/>
            <EvidenceRef feature_ref="ft_73" assay_refs="ass_1"/>
      ...
    </PeptideConsensusList>
```

cvParam Mapping Rules:

```
Path /MzQuantML/PeptideConsensusList
MAY supply a *child* term of MS:1002115 (peptide consensus list attribute) only once
```

6.52 Element <PeptideConsensus_refs>

Definition: The peptides on which the quantitative protein values in the QuantLayer(s) are based. Note this should not be used to report all peptides that can support the protein identification, only quantitation.

Type: xsd:IDREFS

Attributes: none

Subelements: none

Example Context:

```
<PeptideConsensus_refs>pep_LKQSAEEQAQAQAQAQAAAEK pep_LLDAQLSTGGIVDPSK pep_DLLQQFR pep_LSVYTALQR
pep_SIQEELQHLR pep_QLLSPGTALILLEAQAASGFLLDPVR pep_LTVEEAVR pep_YSELTTLTSQYIK pep_SFLSEKDSLLQR
pep_CITDPQTGLCLLPLK pep_YLYGTGAVAGVYLPGSR pep_IDSAEWGVDLPSVEAQLGSHR pep_GYYSPYSVSGSGSTAGSR
pep_LAAISEATR pep_SILDEELQR pep_EVAEADSVR pep_VTLVQTLEIQR pep_TLQEEHVTVAQLR pep_DTHDQLSEPSEVR
pep_AFCGFEDPR pep_LQLEACETR pep_LQAEEVAQQK pep_VYHDPSTQEPVTYSQLQQR pep_APVPASELLDAK pep_SQVEEELFSVR
pep_AQVEQELTTLR pep_HISDLYEDLR pep_WQAVLAQTDVR pep_AQAELEAQELQR pep_AQAEAQQPVFNTLR
pep_SDQLTGLSLLPLSEK pep_LKAEAELLQQQK pep_LAQGHTTVAELTQR pep_SLAAEEEAAR pep_LTVDEAVR pep_LSVAAQEAAR
pep_FLEVQYLTGGLIEPDTPGR pep_ARQEEVYSELQAR pep_EAQAVPATLQELEATK pep_AALAHSEIATTQAASTK
pep_SSSVGSSSSYPISSAGPR pep_DAPDGPSVEAEPEYTFEGLR pep_TQLASWSDPTEETGPVAGILDTETLEK pep_YLQDLLAWVEENQR
pep_GGELVYTDTEAR pep_ALVPAAELLDSGVISHELYQQLQR pep_QLQLAQEAAQK pep_QSAEEQAQAQAQAQAAAEK pep_GLLSAEVAR
pep_LLDAQLATGGIVDPR pep_QEQIQAVPIANCQAAR pep_LQNVQIALDYLR pep_LTVNEAVK pep_CVEDPETGLR pep_NLVDNITGQR
pep_LECLQR pep_VSLDEALQR pep_VLALPEPSPAAPTLR pep_SWSLVTFR pep_VPLDVAYAR pep_QVQVALETAQR pep_LPVDVAYQR
pep_IISLETYNLFR pep_LSYTQLLK pep_GFFDPNTHENLTYLQLLER pep_LAEVEAALEK pep_QITVEELVR pep_ESADPLSAWLQDAK
pep_AKLEQLFQDEVAK pep_VQSGSESVIQEYVDLR pep_DLLQPEVAVALLEAQAGTGHIIDPATSAR pep_AHLLTSGPPPDEQDFIQAYEEVR
pep_ALLEEIER pep_VTQLLER pep_FLEGTSCIAGVFVDATK pep_RQEQIQAVPIANCQAAR pep_SLLAWQSLSR
pep_VIYQSLEGAVQAGQLK pep_LLFNDVQTLK pep_QTNLENLDQAFSVAER pep_LLEAAAQSSK pep_LTAEDLYEAR
pep_LLDPEDVDVPQPDEK pep_SAEVELQSK pep_QLAEEDLAQQR pep_DPYSGQSVSLFQALK pep_QLLEEELAR pep_QLEMSAEAER
pep_LLLWSQR</PeptideConsensus_refs>
```

6.53 Element <PeptideSequence>

Definition: The amino acid sequence of the (poly)peptide. If a substitution modification has been found, the original sequence should be reported. The peptide sequence is mandatory unless this is a PeptideConsensus that has not been assigned to a peptide sequence.

Type: sequence

Attributes: none

Subelements: none

Example Context:

```
<PeptideSequence>DQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEEPGSETSDAK</PeptideSequence>
```

6.54 Element <Person>

Definition: A person's name and contact details. Any additional information such as the address, contact email etc. should be supplied using CV parameters or user parameters.

Type: PersonType

Attributes:

Attribute Name	Data Type	Use	Definition
firstName	xsd:string	optional	The Person's first name.
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
lastName	xsd:string	optional	The Person's last/family name.
midInitials	xsd:string	optional	The Person's middle initial.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.
Affiliation	0	unbounded	The organization a person belongs to.

Example Context:

```
<Person lastName="Mayer" firstName="Gerhard" id="PERS_GM" name="Gerhard Mayer">
    <Affiliation organization_ref="ORG_MPC"/>
</Person>
```

cvParam Mapping Rules:

```
Path /MzQuantML/AuditCollection/Person
SHOULD supply term MS:1000588 (contact URL) one or more times
SHOULD supply term MS:1000587 (contact address) one or more times
SHOULD supply term MS:1000589 (contact email) one or more times
```

6.55 Element <ProcessingMethod>

Definition: Description of one step within the data processing pipeline.

Type: ProcessingMethodType

Attributes:

Attribute Name	Data Type	Use	Definition
order	xsd:positiveInteger	required	This attributes allows a series of consecutive steps to be placed in the correct order, start counting from 1.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<ProcessingMethod order="3">
    <userParam value="feature detection and quantitation" name="Plugin type"/>
```

```
    <userParam value="iTraqQuantitation" name="Plugin name"/>
    <userParam value="examples/paper_iTraq4plex/iTraqQuantitationPSIMOD.xtp" name="Plugin
configuration file"/>
    <userParam value="mean" name="Feature to peptide inference method"/>
    <userParam value="weightedAverage" name="Peptide to protein inference method"/>
    <userParam value="sum" name="Assay to Study Variables inference method"/>
  ...
</ProcessingMethod>
```

cvParam Mapping Rules:

```
Path /MzQuantML/DataProcessingList/DataProcessing/ProcessingMethod
MAY supply a *child* term of MS:1001861 (quantification data processing) one or more times
  e.g.: MS:1001862 (normalization to mean of sum of all proteins)
  e.g.: MS:1001863 (quantile normalization, proteins)
  e.g.: MS:1001864 (quantile normalization, peptides)
  e.g.: MS:1001865 (Progenesis automatic alignment)
  e.g.: MS:1001866 (Progenesis manual alignment)
  e.g.: MS:1001867 (Progenesis normalization)
  e.g.: MS:1002070 (t-test)
  e.g.: MS:1002071 (ANOVA-test)
```

Example cvParams:

```
<cvParam accession="MS:1001865" cvRef="PSI-MS" name="Progenesis automatic alignment"/>
<cvParam accession="MS:1001867" cvRef="PSI-MS" name="Progenesis normalization"/>
<cvParam accession="MS:1001861" cvRef="PSI-MS" name="quantification data processing"/>
<cvParam cvRef="MS" accession="MS:1001862" name="normalization to mean of sum of all proteins"/>
```

Example userParams:

```
<userParam name="Conversion to mzML format" value="ProteoWizard" />
<userParam name="Data filtering" value="FileFilter" />
<userParam name="Data processing action" value="SILACAnalyzer" />
<userParam name="Peak picking" value="SILACAnalyzer" />
<userParam name="Quantitation" value="SILACAnalyzer" />
<userParam value="load identification" name="Plugin type"/>
<userParam value="loadMzIdentML" name="Plugin name"/>
<userParam value="examples/paper_iTraq4plex/loadMzIdentML.xtp" name="Plugin configuration file"/>
<userParam value="mean" name="Feature to peptide inference method"/>
<userParam value="weightedAverage" name="Peptide to protein inference method"/>
<userParam value="sum" name="Assay to Study Variables inference method"/>
<userParam value="true" name="Protein ratio calculation infer from peptide ratio"/>
<userParam name="Identification mapping" value="IDMapper" />
```

6.56 Element <Protein>

Definition: One protein that has been quantified in the file, including references to peptides on which the quantification is based.

Type: ProteinType

Attributes:

Attribute Name	Data Type	Use	Definition
accession	xsd:string	required	The accession of the protein in the source database. In most use cases it is expected that accession will be unique within the ProteinList, although in rare cases there may be different entries for the same protein for example if quantifying different PTMs on the same protein.
id	xsd:ID	required	The unique identifier for the object within the file.
searchDatabase_ref	xsd:IDREF	required	The search database from which this protein was identified

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
IdentificationRef	0	unbounded	Reference for the identification evidence for peptides from the referenced external file and unique identifier e.g. a link to an mzIdentML file and ID for the

			ProteinDetectionHypothesis.
PeptideConsensus_refs	0	1	The peptides on which the quantitative protein values in the QuantLayer(s) are based. Note this should not be used to report all peptides that can support the protein identification, only quantitation.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<Protein searchDatabase_ref="SD1" accession="IPI00421271" id="prot_275">
    <PeptideConsensus_refs>pep_LKQSAEEQAQAQAQAQAAAEK pep_LLDAQLSTGGIVDPSK pep_DLLQQFR pep_LSVYTALQR
pep_SIQEELQHLR pep_QLLSPGTALILLEAQAASGFLLDPVR pep_LTVEEAVR pep_YSELTTLTSQYIK pep_SFLSEKDSLLQR
pep_CITDPQTGLCLLPLK pep_YLYGTGAVAGVYLPGSR pep_IDSAEWGVDLPSVEAQLGSHR pep_GYYSPYSVSGSGSTAGSR
pep_LAAISEATR pep_SILDEELQR pep_EVAEADSVR pep_VTLVQTLEIQR pep_TLQEEHVTVAQLR pep_DTHDQLSEPSEVR
pep_AFCGFEDPR pep_LQLEACETR pep_LQAEEVAQQK pep_VYHDPSTQEPVTYSQLQQR pep_APVPASELLDAK pep_SQVEEELFSVR
pep_AQVEQELTTLR pep_HISDLYEDLR pep_WQAVLAQTDVR pep_AQAELEAQELQR pep_AQAEAQQPVFNTLR
pep_SDQLTGLSLLPLSEK pep_LKAEAELLQQQK pep_LAQGHTTVAELTQR pep_SLAAEEEAAR pep_LTVDEAVR pep_LSVAAQEAAR
pep_FLEVQYLTGGLIEPDTPGR pep_ARQEEVYSELQAR pep_EAQAVPATLQELEATK pep_AALAHSEIATTQAASTK
pep_SSSVGSSSSYPISSAGPR pep_DAPDGPSVEAEPEYTFEGLR pep_TQLASWSDPTEETGPVAGILDTETLEK pep_YLQDLLAWVEENQR
pep_GGELVYTDTEAR pep_ALVPAAELLDSGVISHELYQQLQR pep_QLQLAQEAAQK pep_QSAEEQAQAQAQAQAAAEK pep_GLLSAEVAR
pep_LLDAQLATGGIVDPR pep_QEQIQAVPIANCQAAR pep_LQNVQIALDYLR pep_LTVNEAVK pep_CVEDPETGLR pep_NLVDNITGQR
pep_LECLQR pep_VSLDEALQR pep_VLALPEPSPAAPTLR pep_SWSLVTFR pep_VPLDVAYAR pep_QVQVALETAQR pep_LPVDVAYQR
pep_IISLETYNLFR pep_LSYTQLLK pep_GFFDPNTHENLTYLQLLER pep_LAEVEAALEK pep_QITVEELVR pep_ESADPLSAWLQDAK
pep_AKLEQLFQDEVAK pep_VQSGSESVIQEYVDLR pep_DLLQPEVAVALLEAQAGTGHIIDPATSAR pep_AHLLTSGPPPDEQDFIQAYEEVR
pep_ALLEEIER pep_VTQLLER pep_FLEGTSCIAGVFVDATK pep_RQEQIQAVPIANCQAAR pep_SLLAWQSLSR
pep_VIYQSLEGAVQAGQLK pep_LLFNDVQTLK pep_QTNLENLDQAFSVAER pep_LLEAAAQSSK pep_LTAEDLYEAR
pep_LLDPEDVDVPQPDEK pep_SAEVELQSK pep_QLAEEDLAQQR pep_DPYSGQSVSLFQALK pep_QLLEEELAR pep_QLEMSAEAER
pep_LLLWSQR</PeptideConsensus_refs>
</Protein>
```

cvParam Mapping Rules:

```
Path /MzQuantML/ProteinList/Protein
MAY supply a *child* term of MS:1000884 (protein attribute) one or more times
  e.g.: MS:1000883 (protein short name)
  e.g.: MS:1000885 (protein accession)
  e.g.: MS:1000886 (protein name)
  e.g.: MS:1000933 (protein modifications)
  e.g.: MS:1000934 (gene name)
```

Example userParams:

```
<userParam name="Molecular weight" value="69.1"/>
<userParam name="Number of peptides" value="48"/>
```

6.57 Element <ProteinGroup>

Definition: A grouping of quantified proteins based on ambiguous assignment of peptide evidence to protein identification. The semantics of elements within the group, such as a leading protein or those sharing equal evidence can be reported using cvParams.

Type: ProteinGroupType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
searchDatabase_ref	xsd:IDREF	required	The search database from which this protein group was identified

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
IdentificationRef	0	unbounded	Reference for the identification evidence for peptides from the referenced external file and

			unique identifier e.g. a link to an mzIdentML file and ID for the ProteinDetectionHypothesis.
ProteinRef	1	unbounded	A reference to one of the Proteins contained within this group, along with CV terms describing the role it plays within the group, such as representative or anchor protein, same set or sub-set.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

cvParam Mapping Rules:

```
Path /MzQuantML/ProteinGroupList/ProteinGroup
MAY supply a *child* term of MS:1002113 (protein group attribute) one or more times
```

6.58 Element <ProteinGroupList>

Definition: The list of all groups of proteins with conflicting evidence for which quantitation values are being reported along with quantitative values about those protein groups. If quantitation is done on individual proteins only, ProteinGroupsList should not be included.

Type: ProteinGroupListType

Attributes:

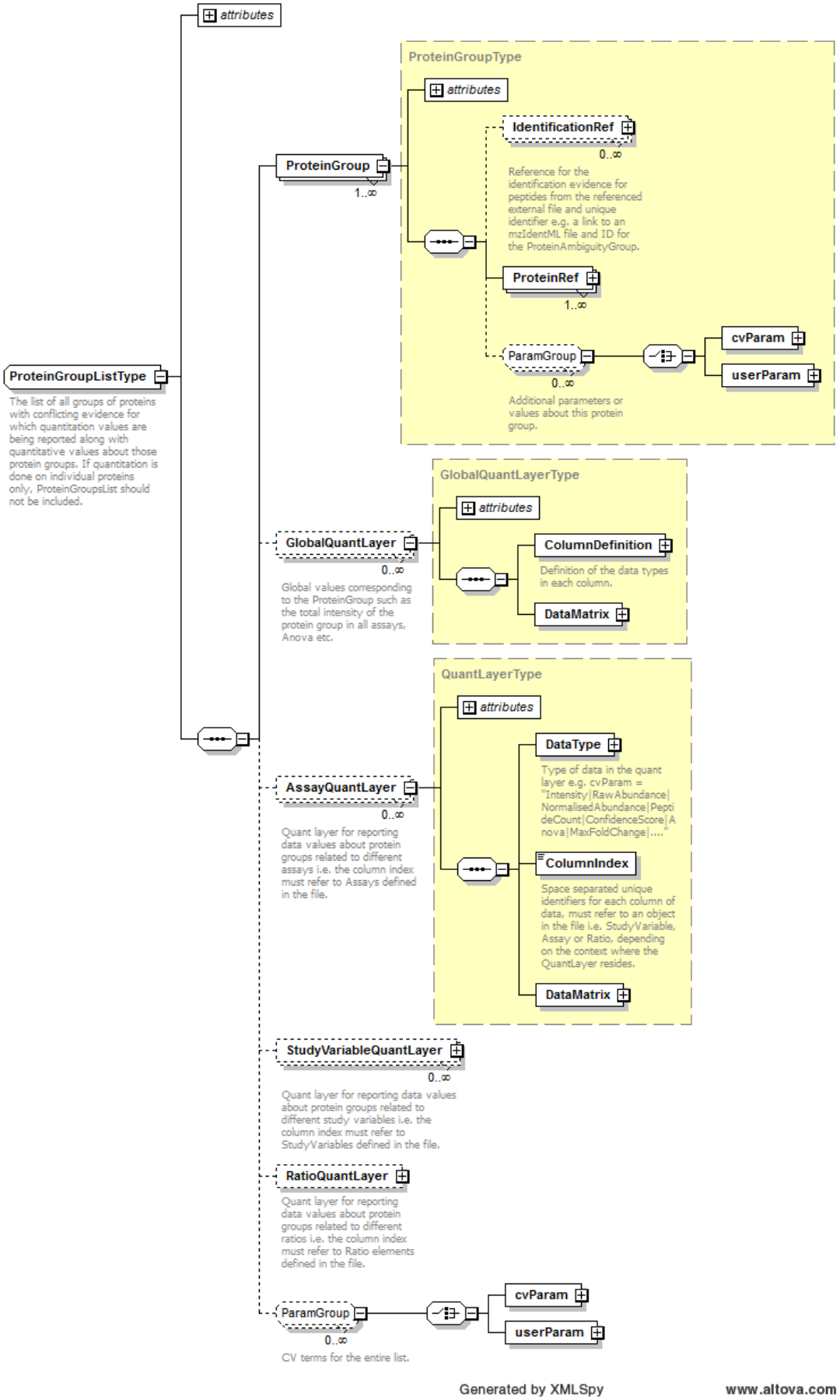
Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ProteinGroup	1	unbounded	A grouping of quantified proteins based on ambiguous assignment of peptide evidence to protein identification. The semantics of elements within the group, such as a leading protein or those sharing equal evidence can be reported using cvParams.
GlobalQuantLayer	0	unbounded	Global values corresponding to the ProteinGroup such as the total intensity of the protein group in all assays, Anova etc.
AssayQuantLayer	0	unbounded	A collection of data relating to the objects within the parent list type (e.g. PeptideConsensus, Protein or ProteinGroup).
StudyVariableQuantLayer	0	unbounded	Quant layer for reporting data values about protein groups related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

RatioQuantLayer	0	1	Quant layer for reporting data values about protein groups related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Graphical Context:



Example Context:

cvParam Mapping Rules:

```
Path /MzQuantML/ProteinGroupList
MAY supply a *child* term of MS:1002112 (protein group list attribute) one or more times
```

6.59 Element <ProteinList>

Definition: The list of all individual proteins (i.e. ungrouped) for which quantitation values are being reported. If quantitation is done on protein groups, the constituent proteins should be listed here with no QuantLayers.

Type: ProteinListType

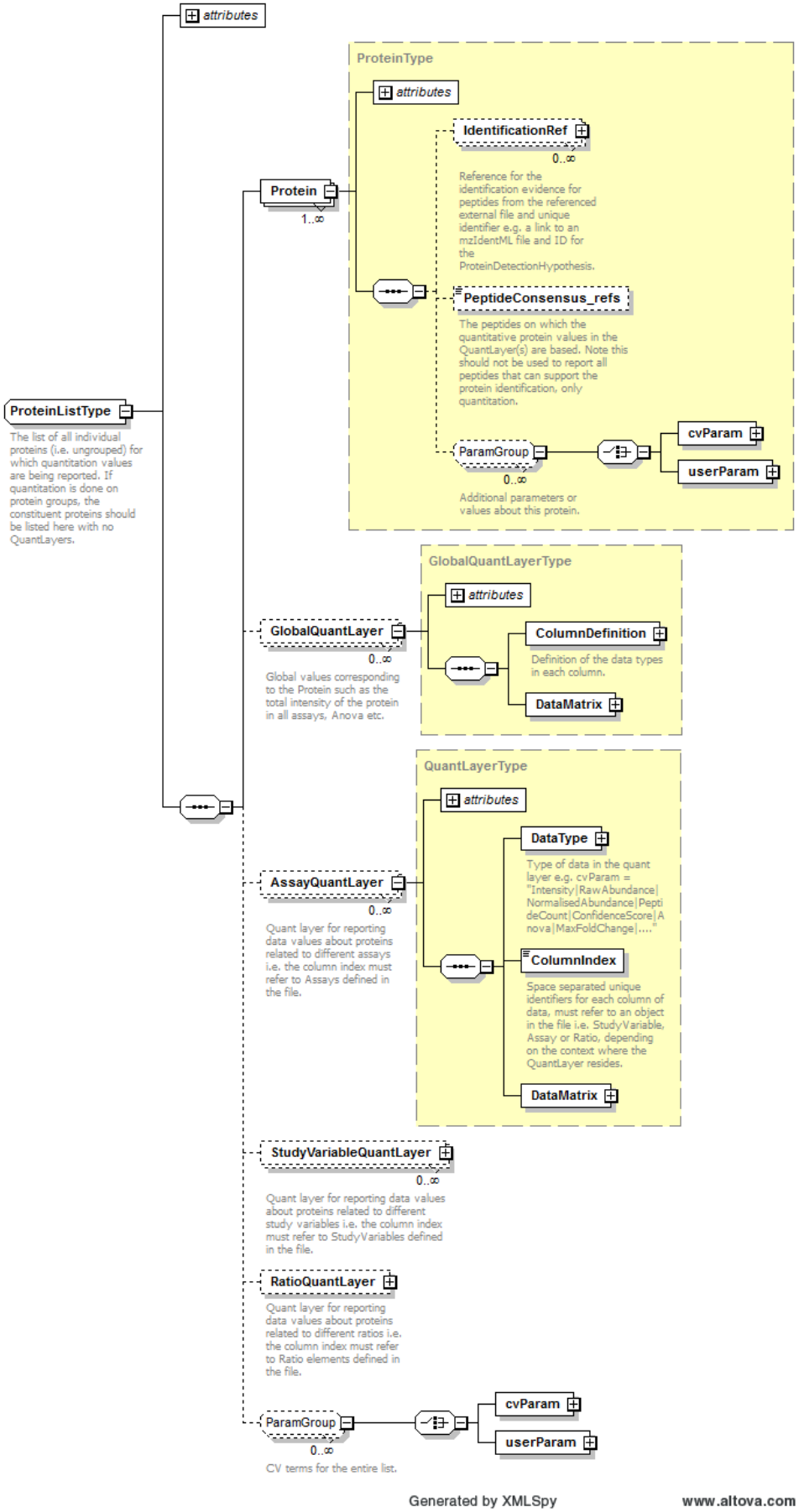
Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Protein	1	unbounded	One protein that has been quantified in the file, including references to peptides on which the quantification is based.
GlobalQuantLayer	0	unbounded	Global values corresponding to the Protein such as the total intensity of the protein in all assays, Anova etc.
AssayQuantLayer	0	unbounded	A collection of data relating to the objects within the parent list type (e.g. PeptideConsensus, Protein or ProteinGroup).
StudyVariableQuantLayer	0	unbounded	Quant layer for reporting data values about proteins related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.
RatioQuantLayer	0	1	Quant layer for reporting data values about proteins related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Graphical Context:



Example Context:

```
<ProteinList id="ProteinList">
    <Protein searchDatabase_ref="SDB_itraq_test_db" accession="sp_P00924_ENO1_YEAST"
id="DBSeq_1_sp_P00924_ENO1_YEAST"/>
    <Protein searchDatabase_ref="SDB_itraq_test_db" accession="sp_P02769_ALBU_BOVIN"
id="DBSeq_1_sp_P02769_ALBU_BOVIN"/>
    <Protein searchDatabase_ref="SDB_itraq_test_db" accession="sp_P62894.2_CYC_BOVIN"
id="DBSeq_1_sp_P62894.2_CYC_BOVIN"/>
    <Protein searchDatabase_ref="SDB_itraq_test_db" accession="sp_P00489_PYGM_RABIT"
id="DBSeq_1_sp_P00489_PYGM_RABIT"/>
    <AssayQuantLayer id="AssayQuantLayer_Proteins_reporter_ion_intensity">
        <DataType>
  ...
</ProteinList>
```

cvParam Mapping Rules:

```
Path /MzQuantML/ProteinList
MAY supply a *child* term of MS:1002114 (protein list attribute) one or more times
```

6.60 Element <ProteinRef>

Definition: A reference to one of the Proteins contained within this group, along with CV terms describing the role it plays within the group, such as representative or anchor protein, same set or sub-set.

Type: ProteinRefType

Attributes:

Attribute Name	Data Type	Use	Definition
protein_ref	xsd:IDREF	required	Reference to one of the proteins within the ProteinList.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

cvParam Mapping Rules:

```
Path /MzQuantML/ProteinGroupList/ProteinGroup/ProteinRef
SHOULD supply a *child* term of MS:1001101 (protein group or subset relationship) one or more
times
  e.g.: MS:1001591 (anchor protein)
  e.g.: MS:1001592 (family member protein)
  e.g.: MS:1001593 (group member with undefined relationship OR ortholog protein)
  e.g.: MS:1001594 (sequence same-set protein)
  e.g.: MS:1001595 (spectrum same-set protein)
  e.g.: MS:1001596 (sequence sub-set protein)
  e.g.: MS:1001597 (spectrum sub-set protein)
  e.g.: MS:1001598 (sequence subsumable protein)
  e.g.: MS:1001599 (spectrum subsumable protein)
```

6.61 Element <Provider>

Definition: The provider of the document in terms of the Contact and the software the produced the document instance.

Type: ProviderType

Attributes:

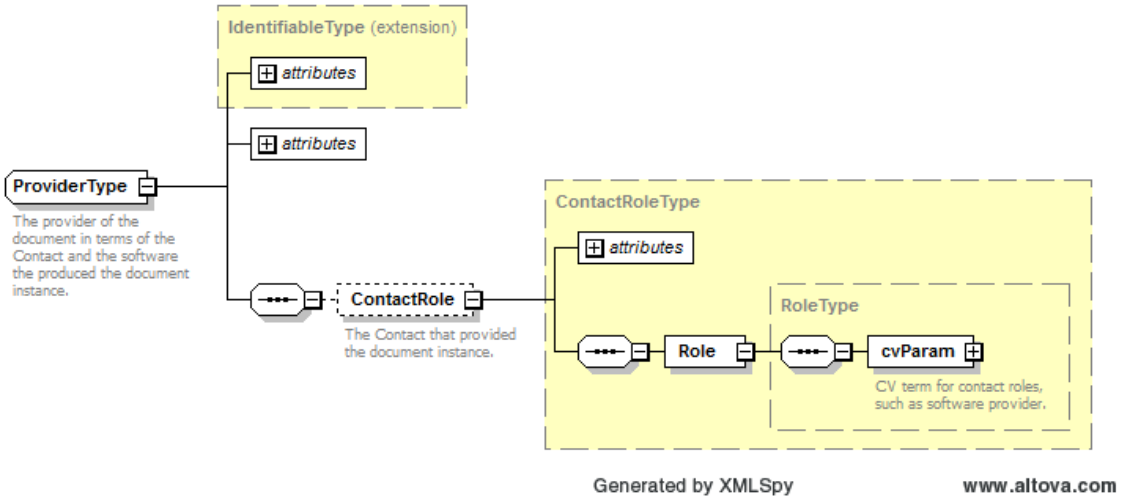
Attribute Name	Data Type	Use	Definition

analysisSoftware_ref	xsd:IDREF	optional	The Software that produced the document instance.
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ContactRole	0	1	The role of the Contact that provided the document instance.

Graphical Context:



Example Context:

```
<Provider analysisSoftware_ref="Microsoft_Excel_2007" id="PROV_1" name="Gerhard Mayer">
    <ContactRole contact_ref="PERS_GM">
        <Role>
            <cvParam cvRef="MS" accession="MS:1001271" name="researcher"/>
        </Role>
    </ContactRole>
</Provider>
```

6.62 Element <Ratio>

Definition: The setup of a ratio of study variables or assays that is referenced elsewhere in the file. It is expected that the numerator and denominator MUST both be Assays or MUST both be StudyVariables. However, StudyVariables MAY contain 1 to many Assays, thus allowing more complex ratios to be constructed if needed via use of StudyVariables with unbalanced numbers of Assays.

Type: RatioType

Attributes:

Attribute Name	Data Type	Use	Definition
denominator_ref	xsd:IDREF	required	Reference to a StudyVariable or an Assay.
id	xsd:ID	required	The unique identifier for the object within the file.
name	xsd:string	optional	Optional name for the ratio.
numerator_ref	xsd:IDREF	required	Reference to a StudyVariable or an Assay.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition

RatioCalculation	0	1	Information should be provided about how ratios are calculated if they differ from simple division of numerator by denominator
NumeratorDataType	1	1	Type of data used for the numerator of the ratio e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFoldChange|...."
DenominatorDataType	1	1	Type of data used for the denominator of the ratio e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance|PeptideCount|ConfidenceScore|Anova|MaxFoldChange|...."

Example Context:

```
<Ratio numerator_ref="ASS_0" denominator_ref="ASS_5" id="RATIO_vIa_30min_vs_vIb_120min"
name="vIa_30min_vs_vIb_120min">
    <NumeratorDataType>
        <cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
    </NumeratorDataType>
    <DenominatorDataType>
        <cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
    </DenominatorDataType>
  ...
</Ratio>
```

6.63 Element <RatioCalculation>

Definition: Information should be provided about how ratios are calculated if they differ from simple division of numerator by denominator

Type: ParamListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	1	unbounded	A single user-defined parameter.

Example Context:

```
<RatioCalculation>
  <userParam name="Simple ratio calc"/>
  <userParam name="light to medium/.../heavy"/>
  <cvParam cvRef="PSI-MS" accession="MS:1001848" name="simple ratio of two values"/>
</RatioCalculation>
```

cvParam Mapping Rules:

```
Path /MzQuantML/RatioList/Ratio/RatioCalculation
SHOULD supply term MS:1001848 (simple ratio of two values) only once
```

Example cvParams:

```
<cvParam accession="MS:1001848" cvRef="PSI-MS" name="simple ratio of two values"/>
```

Example userParams:

```
<userParam name="Simple ratio calc"/>
<userParam name="light to medium/.../heavy"/>
```

6.64 Element <RatioList>

Definition: The definition of ratios of study variables or assays, referenced elsewhere in the document.

Type: RatioListType

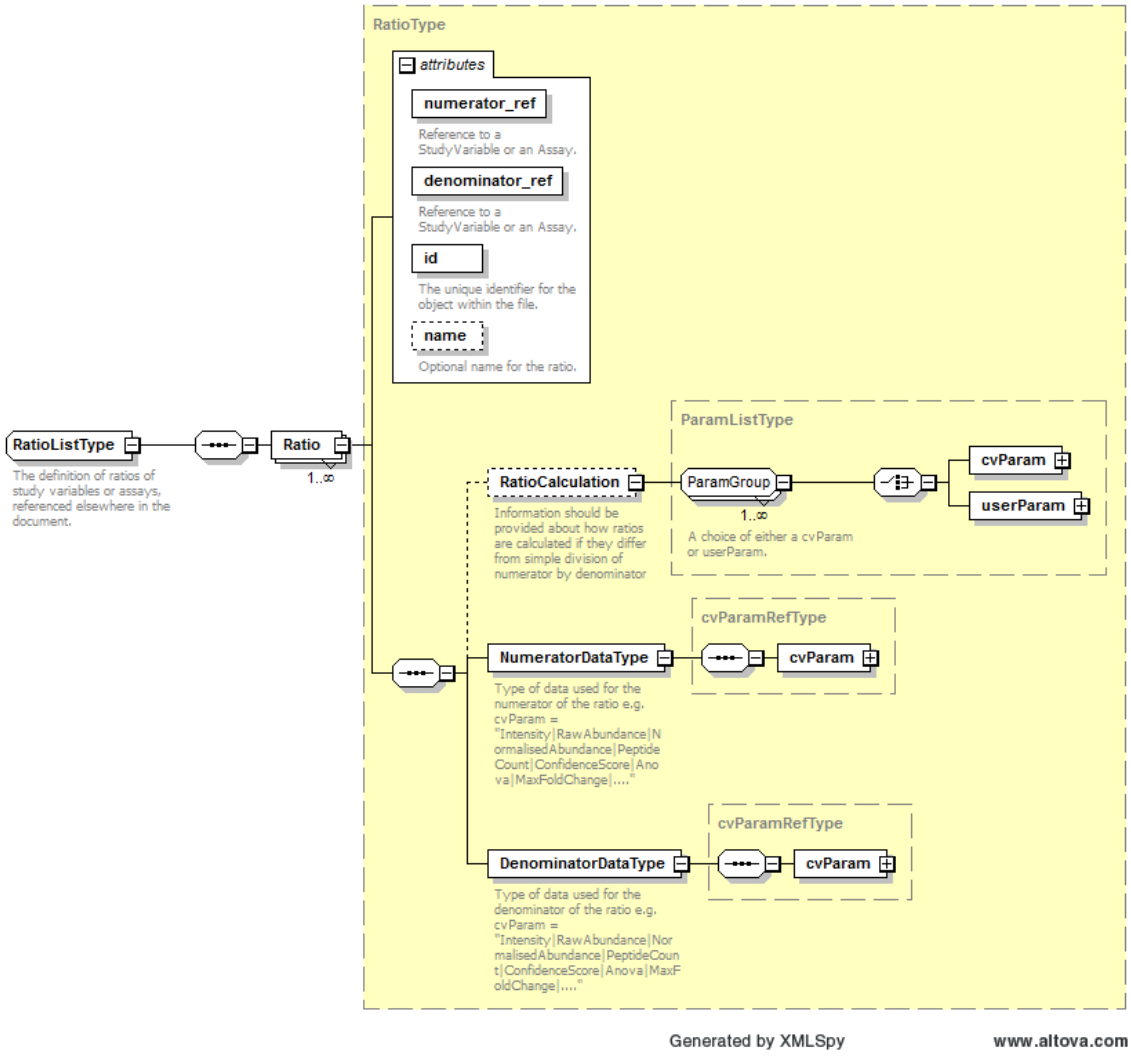
Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition

Ratio	1	unbounded	The setup of a ratio of study variables or assays that is referenced elsewhere in the file. It is expected that the numerator and denominator MUST both be Assays or MUST both be StudyVariables. However, StudyVariables MAY contain 1 to many Assays, thus allowing more complex ratios to be constructed if needed via use of StudyVariables with unbalanced numbers of Assays.

Graphical Context:



Example Context:

```
<RatioList>
    <Ratio numerator_ref="ASS_0" denominator_ref="ASS_5" id="RATIO_vIa_30min_vs_vIb_120min"
name="vIa_30min_vs_vIb_120min">
        <NumeratorDataType>
            <cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
        </NumeratorDataType>
        <DenominatorDataType>
            <cvParam cvRef="MS" accession="MS:1001842" name="peptide PSM count"/>
  ...
</RatioList>
```

6.65 Element <RatioQuantLayer>

Definition:

Depending on context:

1: Quant layer for reporting data values about protein groups related to different ratios i.e. the

column index MUST refer to Ratio elements defined in the file.
2: Quant layer for reporting data values about proteins related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
3: Quant layer for reporting data values about peptides related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
4: Quant layer for reporting data values about small molecules related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.

Type: RatioQuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ColumnIndex	1	1	Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

```
<RatioQuantLayer id="RatioQuantLayer_Proteins">
    <ColumnIndex>ratio_SV_A_B ratio_assay_115_114 ratio_assay_116_114
ratio_assay_117_114</ColumnIndex>
    <DataMatrix>
        <Row object_ref="DBSeq_1_sp_P00924_ENO1_YEAST">1.5688765800335178 0.8774300605025027
0.5904007385417112 0.5794867184130507</Row>
        <Row object_ref="DBSeq_1_sp_P02769_ALBU_BOVIN">0.533639611558188 1.6288166536635666
1.9181070399106477 2.8238352619946046</Row>
        <Row object_ref="DBSeq_1_sp_P62894.2_CYC_BOVIN">0.8750566475401153 1.1763661339620837
1.144083486486904 1.3894599209341105</Row>
        <Row object_ref="DBSeq_1_sp_P00489_PYGM_RABIT">0.9404534810133465 1.099880188336868
1.0193601079613397 1.2766021817677453</Row>
  ...
</RatioQuantLayer>
```

6.66 Element <RawFile>

Definition: A raw mass spectrometry output file that has been analysed e.g. in mzML format. The same raw file can be referenced in multiple assays, for example if it contains multiple samples differentially labelled or tagged. Note, the name raw file does not necessarily imply that the file has not been processed, since in some quant methods, processed peak list formats such as MGF or DTA can be used, which could be referenced here.

Type: RawFileType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
location	xsd:anyURI	required	The location of the data file.
methodFile_ref	xsd:IDREF	optional	An optional reference to a methods file used in association with a raw file, for example a TraML file

			used for SRM analysis.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ExternalFormatDocumentation	0	1	A URI to access documentation and tools to interpret the external format of the ExternalData instance. For example, XML Schema or static libraries (APIs) to access binary formats.
FileFormat	0	1	The format of the ExternalData file, for example "tiff" for image files.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<RawFile location="../msmsdata/mam_042408o_CPTAC_study6_6B011.raw"
    name="mam_042408o_CPTAC_study6_6B011.raw" id="raw_0"/>
```

cvParam Mapping Rules:

```
Path /MzQuantML/InputFiles/RawFilesGroup/RawFile
MAY supply a *child* term of MS:1001817 (raw file attribute) one or more times
  e.g.: MS:1001818 (one sample run)
  e.g.: MS:1001819 (two sample run)
  e.g.: MS:1001820 (three sample run)
  e.g.: MS:1001821 (four sample run)
  e.g.: MS:1001822 (eight sample run)
```

6.67 Element <RawFilesGroup>

Definition: The raw file or collection of raw files that together form one unit of analysis. This is mandatory unless raw files were not used for quantitation e.g. spectral counting. Multiple raw files should only be provided within a group if they have been used for sample pre-fractionation which are later summed together.

Type: RawFilesGroupType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	Unique identifier for the group of raw files that constitute one analysis unit.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
RawFile	1	unbounded	A raw mass spectrometry output file that has been analysed e.g. in mzML format. The same raw file can be referenced in multiple assays, for example if it contains multiple samples differentially labelled or tagged. Note, the name raw file does not necessarily imply that the file has not been processed, since in some quant methods, processed peak list formats

			such as MGF or DTA can be used, which could be referenced here.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<RawFilesGroup id="rg_0">
    <RawFile location="../msmsdata/mam_042408o_CPTAC_study6_6B011.raw"
        name="mam_042408o_CPTAC_study6_6B011.raw" id="raw_0"/>
</RawFilesGroup>
```

cvParam Mapping Rules:

```
Path /MzQuantML/InputFiles/RawFilesGroup
MAY supply a *child* term of MS:1001823 (raw files group attribute) one or more times
  e.g.: MS:1001824 (merge of runs of 1D gel bands)
```

6.68 Element <Role>

Definition: The roles (lab equipment sales, contractor, etc.) the Contact fills.

Type: RoleType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	1	1	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
<Role>
    <cvParam cvRef="MS" accession="MS:1001271" name="researcher"/>
</Role>
```

cvParam Mapping Rules:

```
Path /MzQuantML/Provider/ContactRole/Role
MUST supply a *child* term of MS:1001266 (role type) one or more times
  e.g.: MS:1001267 (software vendor)
  e.g.: MS:1001268 (programmer)
  e.g.: MS:1001269 (instrument vendor)
  e.g.: MS:1001270 (lab personnel)
  e.g.: MS:1001271 (researcher)
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1001271" name="researcher"/>
```

6.69 Element <Row>

Definition: One row of data in a data matrix.

Type: RowType

Attributes:

Attribute Name	Data Type	Use	Definition
object_ref	xsd:IDREF	required	Reference to the data type represented in this row e.g. Feature, Peptide, Protein.

Subelements: none

Example Context:

```
<Row object_ref="raw1-sp_P00489_PYGM_RABIT-WVDTQVVLAMPYDTPVPGYR_00000000000000000000000-3-
817.4252_SIR_2585">0.11366952890220657 0.22005038249250483 1.0851682934290989E-4
0.18373924930756175</Row>
```

6.70 Element <SearchDatabase>

Definition: A database used for searching mass spectra. Examples include a set of amino acid sequence

entries, or annotated spectra libraries.

Type: SearchDatabaseType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
location	xsd:anyURI	required	The location of the data file.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.
numDatabaseEntries	xsd:long	optional	The total number of entries in the database.
releaseDate	xsd:dateTime	optional	The date and time the database was released to the public; omit this attribute when the date and time are unknown or not applicable (e.g. custom databases).
version	xsd:string	optional	The version of the database.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ExternalFormatDocumentation	0	1	A URI to access documentation and tools to interpret the external format of the ExternalData instance. For example, XML Schema or static libraries (APIs) to access binary formats.
FileFormat	0	1	The format of the ExternalData file, for example "tiff" for image files.
DatabaseName	1	1	The database name may be given as a cvParam if it maps exactly to one of the release databases listed in the CV, otherwise a userParam should be used.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.

Example Context:

```
    <SearchDatabase id="sdb_10827256740175179748" location="SwissProtHuman554TD_20080526.fasta">
      <DatabaseName>
        <cvParam cvRef="PSI-MS" accession="MS:1001013" name="database name"
value="SwissProtHuman554TD_20080526.fasta"/>
      </DatabaseName>
    </SearchDatabase>
```

cvParam Mapping Rules:

```
Path /MzQuantML/InputFiles/SearchDatabase
MAY supply a *child* term of MS:1000561 (data file checksum type) one or more times
  e.g.: MS:1000568 (MD5)
  e.g.: MS:1000569 (SHA-1)
```

6.71 Element <SmallMolecule>

Definition: An element to represent a unique identifier of a small molecule for which quantitative values

are reported.

Type: SmallMoleculeType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Modification	0	unbounded	A small molecule modification specification, given by cvParams.
DBIdentificationRef	0	unbounded	External database references for the small molecule identification.
Feature_refs	0	1	Optional references to features on which quantification values about the SmallMolecule in the QuantLayer were based.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

cvParam Mapping Rules:

```
Path /MzQuantML/SmallMoleculeList/SmallMolecule
MAY supply a *child* term of MS:1002118 (small molecule attribute) only once
```

6.72 Element <SmallMoleculeList>

Definition: List of small molecules and associated data values.

Type: SmallMoleculeListType

Attributes:

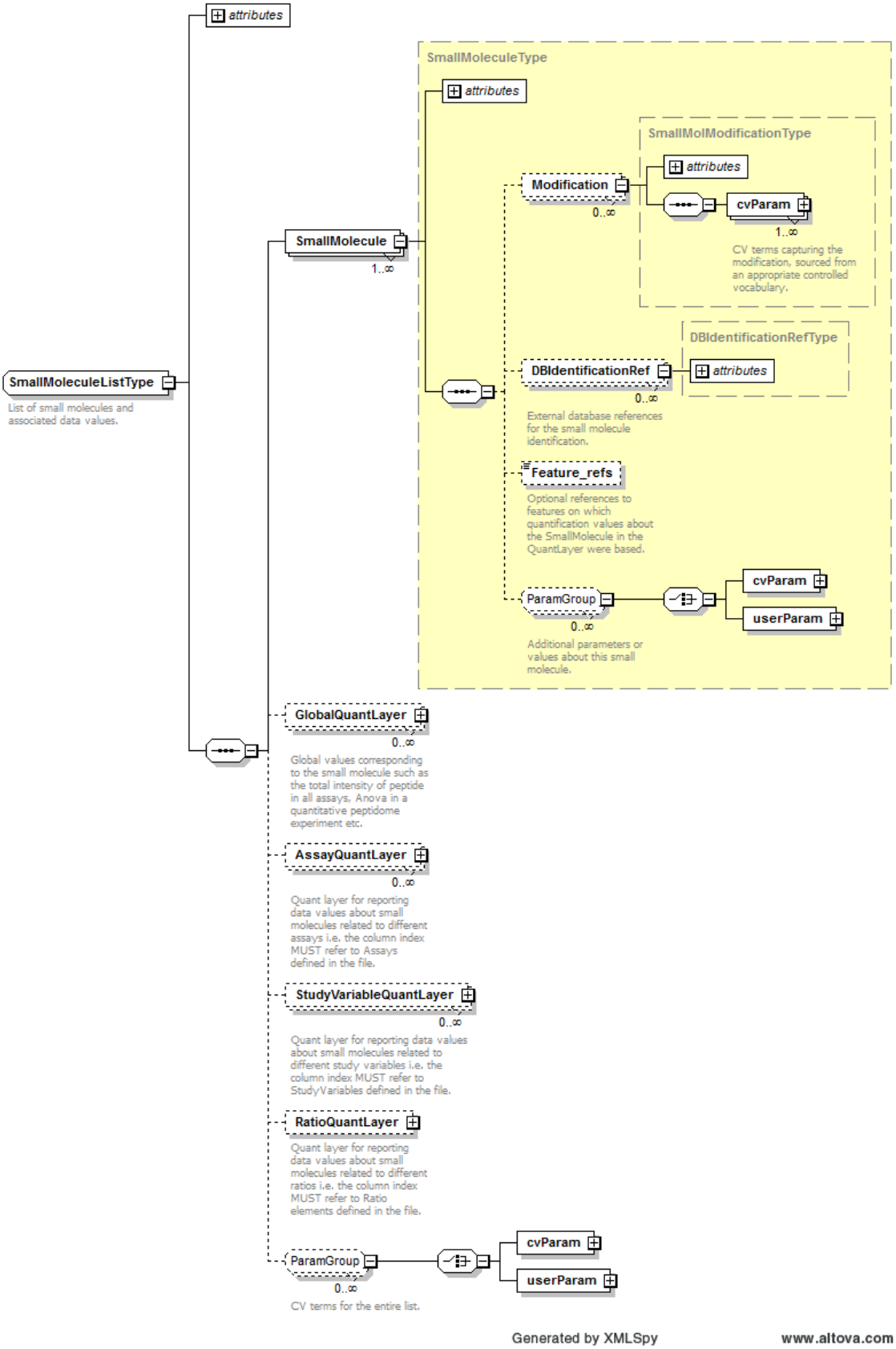
Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
SmallMolecule	1	unbounded	An element to represent a unique identifier of a small molecule for which quantitative values are reported.
GlobalQuantLayer	0	unbounded	Global values corresponding to the small molecule such as the total intensity of the molecule in all assays, Anova etc.
AssayQuantLayer	0	unbounded	Quant layer for reporting data values about small molecules related to different assays i.e. the column index MUST refer to Assays defined in the file.
StudyVariableQuantLayer	0	unbounded	Quant layer for reporting data values about small molecules related to different

			study variables i.e. the column index MUST refer to StudyVariables defined in the file.
RatioQuantLayer	0	1	Quant layer for reporting data values about small molecules related to different ratios i.e. the column index MUST refer to Ratio elements defined in the file.
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Graphical Context:



Example Context:

cvParam Mapping Rules:

```
Path /MzQuantML/SmallMoleculeList
MAY supply a *child* term of MS:1002117 (small molecule list attribute) only once
```

6.73 Element <Software>

Definition: A software package used in the analysis.

Type: SoftwareType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
version	xsd:string	required	The software version.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.

Example Context:

```
<Software id="sw_14624999835377019507" version="1.6.1386">
  <cvParam cvRef="PSI-MS" accession="MS:1000615" name="ProteoWizard"/>
</Software>
```

cvParam Mapping Rules:

```
Path /MzQuantML/SoftwareList/Software
MUST supply a *child* term of MS:1001456 (analysis software) only once
  e.g.: MS:1000532 (Xcalibur)
  e.g.: MS:1000533 (Bioworks)
  e.g.: MS:1000534 (MassLynx)
  e.g.: MS:1000535 (FlexAnalysis)
  e.g.: MS:1000536 (Data Explorer)
  e.g.: MS:1000537 (4700 Explorer)
  e.g.: MS:1000539 (Voyager Biospectrometry Workstation System)
  e.g.: MS:1000551 (Analyst)
  e.g.: MS:1000600 (Proteios)
  e.g.: MS:1000601 (ProteinLynx Global Server)
  et al.
MUST supply a *child* term of MS:1001139 (quantitation software name) only once
  e.g.: MS:1000739 (WARP-LC)
  e.g.: MS:1001488 (Mascot Distiller)
  e.g.: MS:1001583 (MaxQuant)
  e.g.: MS:1001830 (Progenesis LC-MS)
  e.g.: MS:1001831 (SILACAnalyzer)
  e.g.: MS:1001946 (PEAKS Studio)
  e.g.: MS:1001947 (PEAKS Online)
  e.g.: MS:1001948 (PEAKS Node)
  e.g.: MS:1002059 (Microsoft Excel)
  e.g.: MS:1002063 (FindPairs)
  et al.
MAY supply term MS:1001832 (quantitation software comment or customizations) or any of its
children one or more times
```

Example cvParams:

```
<cvParam accession="MS:1001830" cvRef="PSI-MS" name="Progenesis LC-MS"/>
<cvParam accession="MS:1001583" cvRef="PSI-MS" name="MaxQuant"/>
<cvParam cvRef="PSI-MS" accession="MS:1000615" name="ProteoWizard"/>
<cvParam cvRef="PSI-MS" accession="MS:1000757" name="FileFilter"/>
<cvParam cvRef="PSI-MS" accession="MS:1001831" name="SILACAnalyzer"/>
<cvParam accession="MS:1002123" cvRef="PSI-MS" name="x-Tracker"/>
<cvParam cvRef="PSI-MS" accession="MS:1002129" name="ITRAQAnalyzer"/>
<cvParam cvRef="PSI-MS" accession="MS:1002192" name="IDMapper"/>
<cvParam cvRef="MS" accession="MS:1002059" name="Microsoft Excel"/>
```

6.74 Element <SoftwareList>

Definition: List and descriptions of software used to acquire and/or process the data in this file.

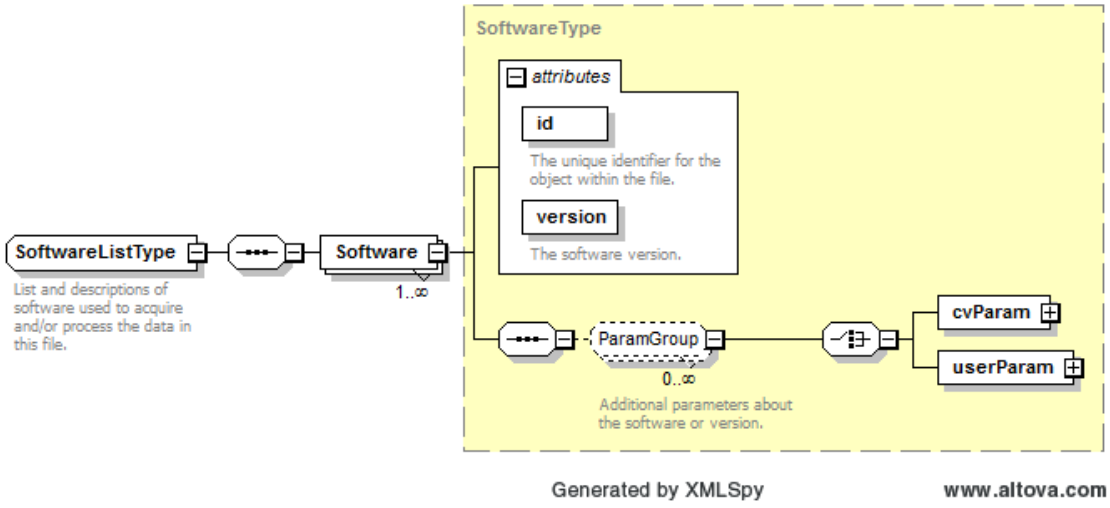
Type: SoftwareListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
Software	1	unbounded	A software package used in the analysis.

Graphical Context:



Example Context:

```
<SoftwareList>
  <Software id="sw_14624999835377019507" version="1.6.1386">
    <cvParam cvRef="PSI-MS" accession="MS:1000615" name="ProteoWizard"/>
  </Software>
  <Software id="sw_4449351221894560547" version="1.7.0">
    <cvParam cvRef="PSI-MS" accession="MS:1000757" name="FileFilter"/>
  </Software>
  ...
</SoftwareList>
```

6.75 Element <SourceFile>

Definition: A file from which this MzQuantML instance was created, including potentially MzQuantML files for earlier stages in a workflow.

Type: SourceFileType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	An identifier is an unambiguous string that is unique within the scope (i.e. a document, a set of related documents, or a repository) of its use.
location	xsd:anyURI	required	The location of the data file.
name	xsd:string	optional	The potentially ambiguous common identifier, such as a human-readable name for the instance.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
ExternalFormatDocumentation	0	1	A URI to access documentation and tools to interpret the external format of the ExternalData instance. For

			example, XML Schema or static libraries (APIs) to access binary formats.
FileFormat	0	1	The format of the ExternalData file, for example "tiff" for image files.

Example Context:

```
<SourceFile location="../SpectralCountInputFiles/" id="SRCF_1" name="counts_prot_per_sample.txt">
    <FileFormat>
        <cvParam cvRef="MS" accession="MS:1000914" name="tab delimited text file"/>
    </FileFormat>
</SourceFile>
```

6.76 Element <StudyVariable>

Definition: A logical grouping of assays into conditions or user-defined study variables such as wild-type versus disease or time points in a time course.

Type: StudyVariableType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.
name	xsd:string	optional	A human readable name for the study variable.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
cvParam	0	unbounded	A single entry from an ontology or a controlled vocabulary.
userParam	0	unbounded	A single user-defined parameter.
Assay_refs	1	1	Reference to the assays that are contained within this study variable.

Example Context:

```
<StudyVariable id="SV_5" name="Ib_120min">
    <cvParam cvRef="MS" accession="MS:1001814" name="generic experimental condition" value="vIb-120min gradient"/>
    <Assay_refs>ASS_5 ASS_6 ASS_7 ASS_8 ASS_9</Assay_refs>
</StudyVariable>
```

cvParam Mapping Rules:

```
Path /MzQuantML/StudyVariableList/StudyVariable
MAY supply a *child* term of MS:1001807 (study variable attribute) one or more times
  e.g.: MS:1001808 (technical replicate)
  e.g.: MS:1001809 (biological replicate)
  e.g.: MS:1001810 (experimental condition 'case')
  e.g.: MS:1001811 (experimental condition 'control')
  e.g.: MS:1001812 (experimental condition 'disease')
  e.g.: MS:1001813 (experimental condition 'healthy')
  e.g.: MS:1001814 (generic experimental condition)
  e.g.: MS:1001815 (time series, time point X)
  e.g.: MS:1001816 (dilution series, concentration X)
```

Example cvParams:

```
<cvParam accession="MS:1001808" cvRef="PSI-MS" name="technical replicate" value=""/>
<cvParam cvRef="MS" accession="MS:1001814" name="generic experimental condition" value="vIa-30min gradient"/>
```

6.77 Element <StudyVariableList>

Definition: The list of experimental conditions used to group results.

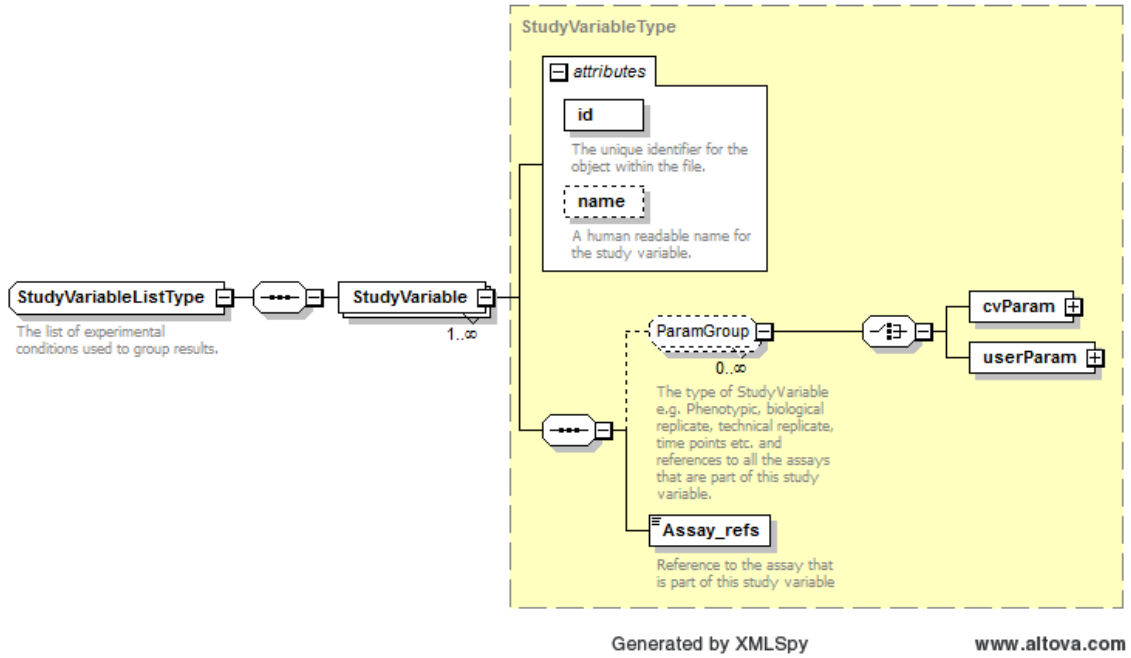
Type: StudyVariableListType

Attributes: none

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
StudyVariable	1	unbounded	A logical grouping of assays into conditions or user-defined study variables such as wild-type versus disease or time points in a time course.

Graphical Context:



Example Context:

```
<StudyVariableList>
    <StudyVariable id="SV_0" name="Ia_30min">
        <cvParam cvRef="MS" accession="MS:1001814" name="generic experimental condition" value="vIa-30min gradient"/>
        <Assay_refs>ASS_0 ASS_1 ASS_2 ASS_3 ASS_4</Assay_refs>
    </StudyVariable>
    <StudyVariable id="SV_5" name="Ib_120min">
        <cvParam cvRef="MS" accession="MS:1001814" name="generic experimental condition" value="vIb-120min gradient"/>
  ...
</StudyVariableList>
```

6.78 Element <StudyVariableQuantLayer>

Definition:

Depending on context:

1: Quant layer for reporting data values about protein groups related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

2: Quant layer for reporting data values about proteins related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

3: Quant layer for reporting data values about peptides related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

4: Quant layer for reporting data values about small molecules related to different study variables i.e. the column index MUST refer to StudyVariables defined in the file.

Type: QuantLayerType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xsd:ID	required	The unique identifier for the object within the file.

Subelements:

Subelement Name	minOccurs	maxOccurs	Definition
DataType	1	1	Type of data in the column e.g. cvParam = "Intensity|RawAbundance|NormalisedAbundance".
ColumnIndex	1	1	Space separated unique identifiers for each column of data, MUST refer to an object in the file i.e. Ratio elements.
DataMatrix	1	1	A matrix of data stored in rows and columns, as defined in the parent QuantLayer.

Example Context:

```
<StudyVariableQuantLayer id="SV_QuantLayer_Protein_reporter_ion_intensity">
    <DataType>
        <cvParam accession="MS:1001847" cvRef="PSI-MS" name="reporter ion intensity"/>
    </DataType>
    <ColumnIndex>SV_group_A SV_group_B</ColumnIndex>
    <DataMatrix>
        <Row object_ref="DBSeq_1_sp_P00924_ENO1_YEAST">12.543762025430688 7.995378467031932</Row>
  ...
</StudyVariableQuantLayer>
```

6.79 Element <userParam>

Definition: A single user-defined parameter.

Type: UserParamType

Attributes:

Attribute Name	Data Type	Use	Definition
name	xsd:string	required	The name of the parameter.
type	xsd:string	optional	The datatype of the parameter, where appropriate (e.g.: xsd:float).
unitAccession	xsd:string	optional	An accession number identifying the unit within the OBO foundry Unit CV.
unitCvRef	xsd:string	optional	If a unit term is referenced, this attribute MUST refer to the CV 'id' attribute defined in the cvList in this file.
unitName	xsd:string	optional	The name of the unit.
value	xsd:string	optional	The user-entered value of the parameter.

Subelements: none

Example Context:

```
<userParam value="examples/paper_iTraq4plex/iTraqQuantitationPSIMOD.xtp" name="Plugin configuration file"/>
```

7. Specific Comments on schema

In this section, several points of documentation are elaborated beyond the core specification in Section 6.

7.1 File extension and compression

It is noted that standard file compression algorithms greatly reduce the mzQuantML file sizes, speeding up file transfers and uploads / downloads. It is also noted that software implementing mzQuantML import or export will be expected to benefit in performance from working with compressed mzQuantML, since the compression and

decompression algorithms are expected to give significant performance gains over disk access times for non-compressed files. As such, it is RECOMMENDED that mzQuantML files are compressed using gzip from all software that exports mzQuantML and software that imports SHOULD be expected to read gzipped files, as well as native (non-compressed) mzQuantML files. The file extension for native mzQuantML files SHOULD be ".mzq" and for compressed files SHOULD be "mzq.gz".

7.2 Referencing elements within the document

A number of elements within the schema have an attribute which is used to reference an element elsewhere in the file using the unique identifier of the referenced element. These attributes are named following the convention: "[elementName]_ref". The uniqueness of the value in the "id" attribute of elements is validated using xsd:ID, and the integrity of the reference is validated using xsd:IDREF, defined within the schema. As such, using XML Schema validation alone, it would be possible to reference an incorrect type of element (since IDREF does not enforce the type of element). However, for a file to be semantically valid, references to the correct type of element MUST be provided and the semantic validation software checks these rules.

7.3 Unknown modifications

In version 1.0.0, only cvParam elements can be given on <PeptideConsensus>:<Modification> and a term "unknown modification" has been added to the PSI-MS CV. This term MUST only be used if the identified modification is not present in UNIMOD (or other allowed CV), according to the identity of the residue modified and the delta mass, within the parent tolerance specified in the search. The semantic validator will check any uses of the "unknown modification" term (MS:1001460) and reject files if the modification is present in UNIMOD.

8. Conclusions

This document contains the specifications for using the mzQuantML format to represent results from peptide and protein identification pipelines, in the context of a proteomics investigation. This specification, in conjunction with the XML Schema, mapping file and CV constitute a proposal for a standard from the Proteomics Standards Initiative. These artefacts are currently undergoing the PSI document process standardization process, which will result in a standard officially sanctioned by PSI.

9. Authors and Contributors

Authors of this specification:

- Mathias Walzer, Center for Bioinformatics/Dept. of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany
- Oliver Kohlbacher, Center for Bioinformatics/Dept. of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany
- Florian Reisinger, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
- J. Alberto Medina-Aunon, Proteomics Facility, Centro Nacional de Biotecnología - CSIC, Madrid, Spain
- Julian Uszkoreit, Medizinisches Proteom-Center, Ruhr-Universität Bochum, Germany
- Gerhard Mayer, Medizinisches Proteom-Center, Ruhr-Universität Bochum, Germany
- Martin Eisenacher, Medizinisches Proteom-Center, Ruhr-Universität Bochum, Germany
- Andrew R Jones, Institute of Integrative Biology, University of Liverpool, UK

Correspondence - andrew.jones@liv.ac.uk

In addition to the authors, numerous people contributed to the model development, gave feedback or tested mzQuantML. A complete list will populate these specifications once the format has completed the PSI document process.

10. References

- [RFC2119] Bradner, S. (1997). "Key words for use in RFCs to Indicate Requirement Levels, Internet Engineering Task Force, RFC 2119, http://www.ietf.org/rfc/rfc2119.txt.
- [Jones 07] Jones AR, Miller M, Spellman P and Pizarro A. Specification documentation for the Functional Genomics Experiment (FuGE) model: user guide. Version 1 (final): http://fuge.sourceforge.net/dev/V1Final/FuGE-v1-SpecDoc.doc.
- [Martens11] Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang W, Rompp A, Neumann S, Pizarro A, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz P, Deutsch E. mzML--a Community Standard for Mass Spectrometry Data MOLECULAR & CELLULAR PROTEOMICS, 10, R110 000133, 2011

11. Intellectual Property Statement

The PSI takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the PSI Chair.

The PSI invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to practice this recommendation. Please address the information to the PSI Chair (see contacts information at PSI website).

Copyright Notice