

1	Overview	1
2	Format specification	1
2.1	Sections	1
2.2	Use of Controlled Vocabulary (CV) terms	2
2.3	Project Metadata	3
2.4	File Mapping	9
2.4.1	<i>File mapping column headers</i>	<i>9</i>
2.4.2	<i>Supported submission file types.....</i>	<i>10</i>
2.5	Sample metadata	11
2.5.1	<i>Sample metadata column headers</i>	<i>11</i>
3	Examples	13
4	References	15

1 Overview

The ProteomeXchange (PX) submission file is a tab-delimited text file, which can be used:

1. For capturing descriptive information about a PX submission (such as essential project metadata, file mappings and sample metadata), to enable automatic processing of the submission.
2. For allowing the handling of 'bulk submissions' (large-scale submissions) and/or integration of the PX submission into semi-automatic pipelines. These are submissions where the high number of files makes the use of the PX submission tool (see below) impractical or where pipelines have been put in place to automatically generate all the information needed for a PX submission.

This file must be submitted together with the other MS-related data files which are part of the submission, and there must only be one PX submission Summary file per dataset (equivalent to one PX accession number).

Although the PX submission tool (available at <http://www.proteomexchange.org/submission>) will generate such a file automatically, there are use cases where the use of a graphical tool is not practical for the submitters. This can happen, for example, when a submission contains hundreds of files, or when a semi-automated submission pipeline is set-up. In those cases, it may be more efficient to generate these submission files using a custom script or program.

For submissions including files of up to a few GBs the use of the PX submission tool is recommended. For larger submissions or automated pipelines, especially if some scripting or programming experience is not available, the PX submission summary file can be exported by the tool and the file transfer be made independently.

In addition, it is also important to highlight that once the PX submission summary file has been created by scripting and all the files that are going to be submitted are available, the PX submission tool can also read the PX summary file and complete the submission (file transfer to the EBI), without having to go through all the steps in the graphical user interface (GUI).

2 Format specification

2.1 Sections

The PX submission file format defines three sections: a section for project metadata followed by a section for the file listing and mappings, then a section for sample metadata. All non-empty lines have to start with one line prefix. For the metadata section the prefix is defined as MTD, whereas the file mapping section has two prefixes. A single line with prefix FMH indicates the header line for this section defining the columns of the following entries. Each of the entry lines has to start with a FME prefix. Also for the sample metadata section, a header line starts with a SMH prefix, and a SME prefix for each of the entry lines. Detailed explanations and examples are given below.

This section describes the structure of a PX submission file.

- **Field separator**
The column delimiter is the Unicode Horizontal Tab character (Unicode codepoint 0009).
- **File encoding**
The UTF-8 encoding of the Unicode character set is the preferred encoding for PX submission files. However, parsers should be able to recognize commonly used encodings.
- **Case sensitivity**
All column labels and field names are case-sensitive.
- **Line prefix**
Every non-empty line in a PX submission file must start with a three-letter code identifying the type of line delimited by a Tab character. The three letter codes are as follows:
 - MTD for metadata lines.
 - FMH for the file mapping header line (containing the column labels).
 - FME for rows of the file mapping section.
 - COM for comment lines.
- **Dates**
If dates and times are used they should be supplied in the ISO 8601 format (“YYYY-MM-DD”, “YYYY-MM-DDTHH:MMZ”, respectively).
- **Decimal separator**
In PX submission files the dot (“.”) MUST be used as decimal separator. Thousand separators MUST NOT be used.
- **Comment lines and empty lines**
Comment lines can be placed anywhere in the file. These lines must start with the three-letter code COM and are ignored by most parsers. Empty lines can also occur anywhere in the file and are ignored.
- **Strings**
Since the format makes use of the Tab character to separate values, attention has to be paid to NOT include any Tab characters in String values. It is best to avoid special characters where not needed. Length limitations may apply, please check the following sections. If no limitation is specifically provided, a default of 500 characters applies.
- **Params**
Certain metadata entries (see Section 2.1 ‘MetaData’) of the PX submission file format make use of controlled vocabulary (CV) terms. For cases where CV terms cannot be given, user params that only contain a name and a value are allowed.
Parameters are always reported as quadruplet [CV label, accession, name, value]. Any field that is not available should be left empty. The length limitation for the ‘value’ field is 200 characters.

```
[NEWT, 9606, Homo sapiens (Human),]
[PSI-MOD, MOD:00696, phosphorylation,]
[,A user parameter, The value]
```

2.2 Use of Controlled Vocabulary (CV) terms

Please note that controlled vocabulary (CV) terms should be used wherever possible, but especially to describe the following metadata: *species*, *tissue*, *cell type*, *disease type*, *instrument (if available in the PSI-MS ontology)* and *protein modifications*. The format of the controlled vocabulary is: [CV Label, CV accession, Name, Value].

- **CV Label** is often the name of the ontology used, for example, for species, NEWT is used

- **CV accession** is the complete accession from the original ontology; for example the NEWT accession for “Homo sapiens (human)” is: “9606”.
- **Name** is the preferred name provided by the ontology; for example the preferred name for human in NEWT is: “Homo sapiens (Human)”.
- **Value** is context specific and may be left empty if not applicable. The length limitation for the ‘value’ field is 200 characters.

Recommended ontologies/ controlled vocabularies:

- for *species*: NEWT (UniProt Taxonomy Database, [1]).
- for *instrument*: MS (PSI Mass Spectrometry CV, [2]).
- for *modification*: either MOD (Protein Modifications PSI-MOD [3]) or Unimod ([4]), with PSI-MOD as the preferred option.
- for *tissues*: BRENDA tissue CV terms.
- for *cell types*: CL cell type CV terms.
- for *diseases*: DOID disease CV terms.

A good place to search for the correct ontology term is the Ontology Lookup Service (OLS, [5]).

2.3 Project Metadata

The metadata section contains entries that generally describe the project. Each line in this section represents an individual metadata entry and is divided into three parts (prefix, type and value). According to the format each line has to start with the line prefix of the metadata section. This is followed by the type of the metadata (Section **Error! Reference source not found.** shows the predefined types), and then the submission specific value of each type. A single tab must exclusively be used to separate the three parts of each line.

submitter_name

Type:	String
Description:	First name and surname of the submitter.
Multiplicity:	1
Example:	MTD submitter_nameJohn Arthur Smith

submitter_email

Type:	String
Description:	Email address of the submitter, this will also be used as the login user name to ProteomeXchange. This is used to assign dataset ownership.
Multiplicity:	1
Example:	MTD submitter_emailjohn@cambridge.ac.uk

submitter_affiliation

Type:	String
Description:	Name of the institute or university which submitter is a member.
Multiplicity:	1
Example:	MTD submitter_affiliation University of Cambridge

lab_head_name

Type:	String
Description:	First name and surname of the principal investigator/lab head. If this person is the same one as the submitter, the information needs to be duplicated.
Multiplicity:	1
Example:	MTD lab_head_nameJohn Arthur Smith

lab_head_email

Type:	String
Description:	Email address of the principal investigator/lab head. If this person is the same one as the submitter, the information needs to be duplicated.
Multiplicity:	1
Example:	MTD lab_head_email john@cambridge.ac.uk

lab_head_affiliation

Type:	String
Description:	Name of the institute or university which the principal investigator/lab head is a member. If this person is the same one as the submitter, the information needs to be duplicated.
Multiplicity:	1
Example:	MTD lab_head_affiliation University of Cambridge

submitter_pride_login

Type:	String
Description:	The PRIDE account user name. Submitters need to have a PRIDE account.
Multiplicity:	1
Example:	MTD submitter_pride_login john@cambridge.ac.uk

project_title

Type:	String
Description:	Title of the project been submitted.
Multiplicity:	1
Example:	MTD project_title Plasma samples from ten male individuals

project_description

Type:	String (minimum 50 characters, maximum 5000 characters)
Description:	A short description of the experiment being submitted. This will be made publicly available in ProteomeCentral and serves as an abstract describing the submission, similar in concept to the abstract of a scientific publication.
Multiplicity:	1
Example:	MTD project_description Here we present a novel tandem IgY12-SuperMix immunoaffinity separation system for enhanced detection of low abundance proteins in human plasma. The tandem IgY12-SuperMix system separates ~60 abundant proteins from the low abundance proteins in plasma, allowing for significant enrichment of low abundance plasma proteins in the SuperMix flow-through fraction. High reproducibility of the tandem separations was observed in terms of both sample processing recovery and LC-MS/MS identification results based on spectral count data.

project_tag

Type:	String
Description:	If the dataset is part of a larger project, this should be indicated here. The project_tags are assigned by the repositories. If you would like to propose a new project_tag, please contact us at: pride-support@ebi.ac.uk
Multiplicity:	0...N
Example:	MTD project_tag Human proteome project MTD project_tag Human plasma project

sample_processing_protocol

Type:	String (minimum 50 characters, maximum 5000 characters)
Description:	A short description of the sample processing protocol being followed.
Multiplicity:	1
Example:	MTD sample_processing_protocol Approximately 40 µg of BMV proteins were resuspended in 8 M deionized urea and reduced and alkylated (TCEP and MMTS, respectively). Proteins were digested for 12 h at 30 °C using 1 µg LysC. The resulting digests were diluted to 2 M urea/20 mM NH ₄ HCO ₃ and digested overnight at °C using 1 µg proteomics grade trypsin. Digested, desalted, and dried peptides were separated using an Agilent 3100 OFFgel fractionator. Samples were separated, concentrated, and analyzed on an Agilent LC-MS system comprised of a 1200 LC system coupled to a 6520 Q-TOF via an HPLC Chip (160 nL, 300 Å C18 150 mm column) Cube interface as described previously (PMID: 21905706).

data_processing_protocol

Type:	String (minimum 50 characters, maximum 5000 characters)
Description:	A short description of the data processing protocol being followed.
Multiplicity:	1
Example:	MTD data_processing_protocol Mass spectrometer RAW data files were converted to MGF format using msconvert from the ProteoWizard distribution. MGF files were searched using X!Hunter against the latest spectral library available on the GPM at the time. X!Tandem and OMSSA searches used Ensembl protein sequence libraries and the annotated HOMD library (274 genomes). All searches included the GPM's cRAP contaminant library. MGF files were searched using X!Tandem using both the native and k-score scoring algorithms and by OMSSA. All searches were performed on Amazon Web Services-based cluster compute instances using Proteome Cluster available at https://www.proteomecluster.com/home . XML output files were parsed and non-redundant protein sets determined using MassSieve.

other_omics_link

Type:	String
Description:	A short string which links to other 'omics' datasets generated by the same study.
Multiplicity:	0..N
Examples:	MTD other_omics_link The A. albimanus transcriptome dataset is available at http://funcgen.vectorbase.org/RNAseq/Anopheles_albimanus/INSP/v2 MTD other_omics_link The corresponding lipidomics study is available at MetaboLights (accession number xxxxx).

keywords

Type:	String
Description:	A list of keywords that describe the content and type of the experiment being submitted. Multiple entries should be comma separated, it is recommended to provide a minimum of three keywords.
Multiplicity:	1
Example:	MTD keywords Human, Plasma, LC-MS

submission_type

Type:	String
Description:	The type of the submission depending on the files to be submitted. Allowed types are: <ul style="list-style-type: none">– COMPLETE: you must provide RAW files and RESULT files as either PRIDE XML or mzIdentML 1.1 format (plus the corresponding spectra in XML-based or

	<p>peak list format).</p> <ul style="list-style-type: none"> – PARTIAL: you must provide RAW files and the corresponding search engine files (e.g. search engine output files, SEARCH files). <p>Please note: we encourage our submitters to go for the COMPLETE submissions, as it represents a full submission to the ProteomeXchange consortium. Once completed, the submitter will be issued with both a ProteomeXchange accession number and a DOI. Also, the dataset from a COMPLETE submission, once made public, will be available in the PRIDE database and made browsable in the PRIDE web page or PRIDE Inspector tool.</p> <p>In contrast, a PARTIAL submission will only be issued with a ProteomeXchange accession, and the dataset will only be available on the public FTP. No DOI will be assigned.</p> <p>Note: in both cases it is possible to provide additional files, for example quantitation results, gel images, etc.</p>
Multiplicity:	1
Example:	MTD submission_type COMPLETE

experiment_type

Type:	CV param
Description:	<p>The current types of the experiment types are listed as CV params below:</p> <p>[PRIDE, PRIDE:0000427, Top-down proteomics,]</p> <p>[PRIDE, PRIDE:0000429, Shotgun proteomics,]</p> <p>[PRIDE, PRIDE:0000430, Chemical cross-linking coupled with mass spectrometry proteomics,]</p> <p>[PRIDE, PRIDE:0000433, Affinity purification coupled with mass spectrometry proteomics,]</p> <p>[PRIDE, PRIDE:0000311, SRM/MRM,]</p>
Multiplicity:	1..N
Example:	MTD experiment_type [PRIDE, PRIDE:0000427, Top-down proteomics,]

reason_for_partial

Type:	String
Description:	Comments to describe why a COMPLETE submission was not possible. Here the submitter must mention the search engine (and/or pipeline) used to generate the results.
Multiplicity:	0..1 (optional for PARTIAL submission, not required for other types of submissions)
Example:	MTD reason_for_partial The file was produced by pipeline XXX using the tools YYY and ZZZ and the output format FFF was not supported by the PRIDE Converter.

species

Type:	CV param
Description:	Controlled vocabulary term to describe a single species. NEWT CV terms are allowed.
Multiplicity:	1..N
Example:	MTD species [NEWT, 9606, Homo Sapiens (Human),]

tissue

Type:	CV param
Description:	Controlled vocabulary term to describe a single tissue. BRENDA Tissue CV terms are allowed [6]. If no tissue is not available, the following CV param should be used instead: [PRIDE, PRIDE: 0000442, Tissue not applicable to dataset,]
Multiplicity:	1..N
Example:	MTD tissue [BRENDA, BTO:0000142, brain,]

cell_type

Type:	CV param
Description:	Controlled vocabulary term to describe a single cell type. Cell Type ontology CV terms are allowed [7].
Multiplicity:	0..N
Example:	MTD cell_type [CL, CL:0000236, B cell,]

disease

Type:	CV param
Description:	Controlled vocabulary term to describe a disease. Disease ontology CV terms are allowed [8].
Multiplicity:	0..N
Example:	MTD disease [DOID, DOID:1319, brain cancer,]

quantification

Type:	CV param
Description:	<p>Controlled vocabulary terms to describe a the quantification technique used in the experiment. At present, PRIDE ontology CV terms are recommended, below are the options:</p> <p>[PRIDE, PRIDE:0000318, 18O,] [PRIDE, PRIDE:0000320, AQUA,] [PRIDE, PRIDE:0000319, ICAT,] [PRIDE, PRIDE:0000321, ICPL,] [PRIDE, PRIDE:0000315, SILAC,] [PRIDE, PRIDE:0000314, TMT,] [PRIDE, PRIDE:0000313, iTRAQ,] [PRIDE, PRIDE:0000323, TIC,] [PRIDE, PRIDE:0000322, emPAI,] [PRIDE, PRIDE:0000435, Peptide counting,] [PRIDE, PRIDE:0000436, Spectral counting,] [PRIDE, PRIDE:0000437, Protein Abundance Index – PAI,] [PRIDE, PRIDE:0000438, Spectrum count/molecular weight,] [PRIDE, PRIDE:0000439, Spectral Abundance Factor – SAF,] [PRIDE, PRIDE:0000440, Normalized Spectral Abundance Factor – NSAF,] [PRIDE, PRIDE:0000441, APEX - Absolute Protein Expression,]</p>
Multiplicity:	0..N
Example:	MTD quantification [PRIDE, PRIDE:0000318, 18O,]

instrument

Type:	CV param
Description:	<p>Controlled vocabulary terms to describe a single instrument model. PSI MS ontology CV terms are allowed.</p> <p>In case the instrument model is not included in the ontology, then use the generic instrument model CV term, [MS, MS:1000031, instrument model, CUSTOM NAME], and replace 'CUSTOM NAME' with your instrument model name.</p>
Multiplicity:	1..N
Example:	MTD instrument [MS, MS:1000447, LTQ,]

modification

Type:	CV param
Description:	Controlled vocabulary term to describe a single modification. PSI-MOD and Unimod CV terms are allowed. Note: if no modification was reported in the experiment, the

	CV term [PRIDE, PRIDE:0000398, No PTMs are included in the dataset,] must be used.
Multiplicity:	1..N (optional for complete submission, required for partial submission)
Example:	MTD modification [MOD,MOD:00394,acetylated residue,]

additional

Type:	CV param or user param
Description:	Additional params that describe the submission. Only a single controlled vocabulary is allowed per line.
Multiplicity:	0..N (optional)
Example:	MTD additional [,,Patient,Colorectal cancer patient 1]

pubmed_id

Type:	String
Description:	The PubMed IDs assigned to the publication related to the dataset (if known). Only a single ID is allowed per line.
Multiplicity:	0..N (optional)
Example:	MTD pubmed_id 18632595

resubmission_px

Type:	String
Description:	If you are resubmitting a previously submitted ProteomeXchange submission, you should reference the original ProteomeXchange accession number. Only one accession is allowed since resubmissions should only happen on a <i>per</i> dataset basis.
Multiplicity:	0..1
Example:	MTD resubmission_px PXD000001

reanalysis_px

Type:	String
Description:	If your experiment is a reanalysis of (one or more) previous PX datasets, these should be referenced here. Multiple entries are allowed, each referencing one PX accession.
Multiplicity:	0..N
Example:	MTD reanalysis_px PXD000001

2.4 File Mapping

The file section details the data files to be submitted and their relationships. Lines in this section are not restricted to three parts, but made up from a number of columns separated from each other with a single tab. The layout closely resembles a table. The start of the section is a header line, which defines the content of each column (and provides the column name). The header line starts with the prefix FMH at the beginning of the line and is followed by the predefined columns listed in Section 2.4.1. The rest of the section is composed of the actual file entries. Each line represents one file and its mappings.

Please note that only predefined file types are allowed (see Section 2.4.2).

Please also note that for a successful submission, **all file names must be unique**.

2.4.1 File mapping column headers

file_id

Type:	String
Description:	Unique file ID, as consecutive non-negative integer starting with 1.

Multiplicity:	1
----------------------	---

file_type

Type:	String
Description:	Predefined file type, see Section 2.4.2.
Multiplicity:	1

file_path

Type:	String
Description:	Absolute path of the file.
Multiplicity:	1

file_mapping

Type:	String
Description:	A list of all the files related to this file, represented using their file ID. For COMPLETE submissions, RESULT files must list the RAW files on which they are based. For PARTIAL submissions, SEARCH files must list the RAW files on which they are based. In all cases, additional mappings to other types of files can be added.
Multiplicity:	1 (required for the COMPLETE case)

2.4.2 Supported submission file types

result

Type:	String
Description:	Result files in either in the formats mzIdentML version 1.1 or in PRIDE XML 2.1. In the case of mzIdentML 1.1, files PEAK will also be mandatory. These PEAK files should be the ones directly used in the search.
Multiplicity:	1..N (required for the COMPLETE case)

raw

Type:	String
Description:	mass spectrometer output files. Binary files, mzML, mzXML and mzData files are allowed. Peak lists files are not allowed, they are represented with a separate type (PEAK).
Multiplicity:	1..N

peak

Type:	String
Description:	Processed peak list file, such as MGF, dta, ms2 and pkl. They could also be heavily processed mzML, mzXML or mzData files. They are optional unless a COMPLETE submission using mzIdentML files are used.
Multiplicity:	0..N (required for COMPLETE submission with mzIdentML files)

search

Type:	String
Description:	Search engine output files, such as Mascot DAT or other output files (from analysis pipelines, such as pep.xml or prot.xml).
Multiplicity:	0..N (required for PARTIAL submissions)

quantification

Type:	String
Description:	Quantification analysis output files, commonly these are the files produced by

	quantification software, such as MaxQuant.
Multiplicity:	0..N (optional)

gel

Type:	String
Description:	Gel image files, commonly these image files are in some kind of image format such as TIF, JPG, PNG, GIF and etc
Multiplicity:	0..N (optional)

other

Type:	String
Description:	Any other files that may be important for the submission, such as scripts, etc
Multiplicity:	0..N (optional)

2.5 Sample metadata

The sample metadata section details the sample information for each RESULT file mentioned in the 'File Mapping' section. Lines in this section are not restricted to three parts, but made up from a number of columns separated from each other with a single tab. The layout closely resembles a table. The start of the section is a header line, which defines the content of each column (and provides the column name). The header line starts with the prefix SMH at the beginning of the line and is followed by the predefined columns listed in Section 2.5.1. The rest of the section is composed of the RESULT files entries. Each line represents one RESULT file and its corresponding sample details.

2.5.1 Sample metadata column headers

file_id

Type:	String
Description:	Unique file ID, as consecutive non-negative integer starting with 1.
Multiplicity:	1

species

Type:	CV param
Description:	Controlled vocabulary term to describe a single species. NEWT CV terms are allowed.
Multiplicity:	1..N
Example:	[NEWT, 9606, Homo sapiens (Human),]

tissue

Type:	CV param
Description:	Controlled vocabulary term to describe a single tissue. BRENDA Tissue CV terms are allowed [6].
Multiplicity:	1..N
Example:	[BRENDA, BTO:0000142, brain,]

cell_type

Type:	CV param
Description:	Controlled vocabulary term to describe a single cell type. Cell Type ontology CV terms are allowed [7].
Multiplicity:	0..N
Example:	[CL, CL:0000236, B cell,]

disease

Type:	CV param
Description:	Controlled vocabulary term to describe a disease. Disease ontology CV terms are allowed [8].
Multiplicity:	0..N
Example:	[DOID, DOID:1319, brain cancer,]

quantification

Type:	CV param
Description:	<p>Controlled vocabulary terms to describe a single quantification technique. PRIDE ontology CV terms are recommended, below are the options:</p> <p>[PRIDE, PRIDE:0000318, 18O,]</p> <p>[PRIDE, PRIDE:0000320, AQUA,]</p> <p>[PRIDE, PRIDE:0000319, ICAT,]</p> <p>[PRIDE, PRIDE:0000321, ICPL,]</p> <p>[PRIDE, PRIDE:0000315, SILAC,]</p> <p>[PRIDE, PRIDE:0000314, TMT,]</p> <p>[PRIDE, PRIDE:0000313, iTRAQ,]</p> <p>[PRIDE, PRIDE:0000323, TIC,]</p> <p>[PRIDE, PRIDE:0000322, emPAI,]</p> <p>[PRIDE, PRIDE:0000435, Peptide counting,]</p> <p>[PRIDE, PRIDE:0000436, Spectrum counting,]</p> <p>[PRIDE, PRIDE:0000437, Protein Abundance Index – PAI,]</p> <p>[PRIDE, PRIDE:0000438, Spectrum count/molecular weight,]</p> <p>[PRIDE, PRIDE:0000439, Spectral Abundance Factor – SAF,]</p> <p>[PRIDE, PRIDE:0000440, Normalized Spectral Abundance Factor – NSAF,]</p> <p>[PRIDE, PRIDE:0000441, APEX - Absolute Protein Expression,]</p>
Multiplicity:	0..N
Example:	[PRIDE, PRIDE:0000318, 18O,]

instrument

Type:	CV param
Description:	<p>Controlled vocabulary terms to describe a single instrument model. PSI MS ontology CV terms are allowed.</p> <p>In case the instrument model is not included in the ontology, then use the generic instrument model CV term, [MS, MS:1000031, instrument model, CUSTOM NAME], and replace 'CUSTOM NAME' with your instrument model name.</p>
Multiplicity:	1..N
Example:	[MS, MS:1000447, LTQ,]

modification

Type:	CV param
--------------	----------

Description:	Controlled vocabulary term to describe a single modification. PSI-MOD and Unimod CV terms are allowed. Note: if no modification was reported in the experiment, the CV term [PRIDE, PRIDE:0000398, No PTMs are included in the dataset,] must be used.
Multiplicity:	0..N
Example:	[MOD,MOD:00394,acetylated residue,]

experimental_factor

Type:	String
Description:	A short description of the experimental factor of the different RESULT files. This information needs to be provided for each RESULT file (linked to the other types of files).
Multiplicity:	1
Examples:	Drug A at 10 mM Drug A at 10 mM, Drug B at 30 mM Wild type strain Technical replicate 2 Experiment done in mice 28 days old (Biological replicate 1) Experiment done in mice 28 days old (Biological replicate 2) Experiment done in mice 10 days old (Biological replicate 2) Disease A Disease B

3 Examples

```

MTD submitter_name John Arthur Smith
MTD submitter_email john.smith@cam.edu
MTD submitter_affiliation University of Cambridge
MTD submitter_pride_login john.smith@cam.edu
MTD lab_head_name Alice Wonderland
MTD lab_head_email alice.wonderland@cam.edu
MTD lab_head_affiliation University of Cambridge
MTD project_title Human proteome
MTD project_description An experiment about human proteome
MTD project_tag Human plasma project
MTD sample_processing_protocol sample processing protocol
MTD data_processing_protocol data processing protocol
MTD other_omics_link http://www.funcgen.vectorbase.org/INSP
MTD keywords human, proteome
MTD submission_type COMPLETE
MTD experiment_type [PRIDE, PRIDE:0000429, Shotgun proteomics, ]
MTD species [NEWT, 9606, Homo Sapiens (Human),]
MTD tissue [BRENDA, BTO:0000142, brain,]
MTD instrument [MS, MS:1000447, LTQ,]
MTD modification [MOD,MOD:00394,acetylated residue,]

FMH file_id file_type file_path file_mapping
FME 1 result /path/to/mzidentml/files/pride-1.xml 2,3,4,5,6,7
FME 2 raw /path/to/raw/files/raw-1.bin
FME 3 raw /path/to/raw/files/raw-2.bin
FME 4 raw /path/to/raw/files/raw-3.bin
FME 5 search /path/to/search/engine/output/search-1.out
FME 6 other /path/to/other/file/other-1.e
FME 7 raw /path/to/peak/list/mzml-1.xml

SMH file_id species tissue instrument modification experimental_factor
SME 1 [NEWT, 9606, Homo sapiens (Human),] [BRENDA, BTO:0000142, brain,]
[MS, MS:1000447, LTQ,] [MOD,MOD:00394,acetylated residue,] Drug A at
10 mM

```

Figure 1: Example file for COMPLETE submission.

```

MTD submitter_name John Arthur Smith
MTD submitter_email john.smith@cam.edu
MTD submitter_affiliation University of Cambridge
MTD submitter_pride_login john.smith@cam.edu
MTD lab_head_name Alice Wonderland
MTD lab_head_email alice.wonderland@cam.edu
MTD lab_head_affiliation University of Cambridge
MTD project_title Human proteome
MTD project_description An experiment about human proteome
MTD project_tag Human plasma project
MTD sample_processing_protocol sample processing protocol
MTD data_processing_protocol data processing protocol
MTD other_omics_link http://www.funcgen.vectorbase.org/INSP
MTD keywords human, proteome
MTD submission_type PARTIAL
MTD experiment_type [PRIDE, PRIDE:0000429, Shotgun proteomics, ]
MTD reason_for_partial cannot convert using the PRIDE Converter
MTD species [NEWT, 9606, Homo Sapiens (Human),]
MTD species [NEWT, 741158, Homo sp. Altai,]
MTD tissue [BRENDA, BTO:0000142, brain,]
MTD instrument [MS, MS:1000447, LTQ,]
MTD modification [MOD,MOD:00394,acetylated residue,]

FMH file_id file_type file_path file_mapping
FME 1 search /path/to/mzidentml/files/mzidentml-1.xml 2,3,4,5,6
FME 2 raw /path/to/raw/files/raw-1.bin
FME 3 raw /path/to/raw/files/raw-2.bin
FME 4 raw /path/to/raw/files/raw-3.bin
FME 5 other /path/to/other/file/other-1.e
FME 6 raw /path/to/peak/list/mzml-1.xml

```

Figure 2: Example file for PARTIAL submission.

4 References

- [1] <http://www.uniprot.org/help/taxonomy>
- [2] <http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo>
- [3] <http://psidev.cvs.sourceforge.net/psidev/psi/mod/data/PSI-MOD.obo>
- [4] <http://www.unimod.org/>
- [5] <http://www.ebi.ac.uk/ontology-lookup/>
- [6] <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=BTO>
- [7] <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=CL>
- [8] <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=DOID>