**mzQC: report and exchange format for mass spectrometry related quality control.**

Status of This Document

This document presents a draft specification of the mzQC data format developed by members of the Human Proteome Organisation (HUPO) Proteomics Standards Initiative (PSI) Quality Control (QC) Working Group. Distribution is unlimited.

Version of This Document

The current version of this document is 0.1.0, November, 2019.

# 

# 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 Quality Control Working Group is developing standards and recommendations for describing the quality of mass spectrometry data and related analysis results. This document defines a JSON file format to report “quality metrics” on a mass spectrometry injection, derived analysis results, or collections thereof. The metrics are defined in the working group’s controlled vocabulary. The metric values are to assist general quality control, (automated) decision making, and visualisation efforts.

**Contents**

Abstract 1

1. Short Summary 4

2. Introduction 4

2.1 Background 4

2.2 Document Structure 5

3. Use Cases for mzQC 5

4. Notational Conventions 6

5. Relationship to Other Specifications 6

5.1 The PSI Mass Spectrometry Controlled Vocabulary (MSCV) 6

5.2 The Quality Control Controlled Vocabulary (QCCV) 7

5.3 MzML 7

5.4 MzIdentML 7

6. Format specification 8

6.1 Schema sections 8

6.1.1 runQuality 8

6.1.2 controlledVocabularies 9

6.1.3 runQualities 9

6.1.4 setQualities 9

6.1.5 qualityMetric 10

6.1.6 setQuality 10

6.1.7 cvParameters 11

6.1.8 inputFiles 11

6.1.9 analysisSoftware 12

6.1.10 metadata 12

6.1.11 cvParameter 13

6.1.12 controlledVocabulary 13

6.1.13 fileFormat 14

6.1.14 fileProperties 14

6.1.15 inputFile 15

6.1.16 qualityMetrics 16

6.1.17 baseQuality 16

6.1.18 mzQC 16

6.2 CvParameter Values for metrics 17

7. Conclusions 17

8. Authors 19

9. Contributors 19

10. References 19

11. Appendix I: qcML to mzQC conversion 21

11.1 Incompatible metrics 21

12. Intellectual Property Statement 22

TradeMark Section 22

Copyright Notice 22

# Short Summary

The mzQC specification describes a file format to report and exchange quality related information about mass spectrometry-based experiments and subsequent analysis. Data are represented in JavaScript Object Notation (JSON) structures and organised in hierarchical fashion, connecting required contextual meta data with quality metric data. Where applicable, such metrics and meta data are defined by a controlled vocabulary to allow for flexibility and updates without changing the specification.

# Introduction

## Background

This document addresses the systematic description of quality control information from mass spectrometry (MS)-based experiments in a text-based format derived from JSON. As the acronym in the name suggests, JSON: JavaScript Object Notation – is handling the same basic data types as a standard JavaScript object: strings, numbers, arrays, booleans, and other object literals – making it possible to build hierarchical schema specifications, like in XML.

XML has been used extensively for describing PSI formats [MZML, MZIDENTLM, MZQUANTML], showing with few specifications for many different kinds of MS-based experiment data, the advantages of flexible serialisation of data. The weaknesses of XML are the added formatting overhead which can lead to relatively large files and the complexity of implementing these formats with available XML handling libraries into new and existing analysis software systems.

However, unlike XML, JSON formats have a small footprint by default, and a wider built-in support to programming languages or development frameworks (libraries of prebuilt tools and structures). XML values are strings of characters, with no built-in type safety, JSON has strong typing built in. This reduces implementation complexity since much work is taken over by the JSON libraries.

XML allows for references instead of duplication of data, introducing further complexity which results in a heavy implementation overhead. Referencing capability in JSON is minimal and must be defined in the format specification. For the purpose of describing quality control information, data can be organised in such way, that the need for referencing is minimised. However external references are needed and will be described in the respective section.

Therefore, JSON can be used for the same kind of data interchange purposes as XML. By specifying mzQC in JSON, we can leverage the broader availability of efficient libraries and simpler implementation while adhering to the proven design principles of PSI formats and maintaining the necessary level of compatibility.

JSON is often lauded "self-describing" and easy to understand, an argument explaining the emergence of JSON as a replacement for XML in many systems. It is the de-facto language of the API, where easy-to-understand, lightweight data-interchange formats are the ones most frequently adopted.

(2)

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. It is anticipated that tutorial material will be developed independently of this specification.

## Document Structure

The remainder of this document is structured as follows. Section 3 lists use cases for the format. Section 4 is devoted to Notational Conventions throughout the document. Section 5 outlines the relationships between mzQC and other file format specifications. Section 6 includes the details of the format specification, listing the required and optional fields. Section 6 is a brief summary of the conclusions. Sections 8, 9 and 10 are devoted to the list of Authors, contributors and references, respectively. There is one Appendix (section 11) devoted to show how qcML files, the progenitor format of mzQC, can be converted to mzQC.

# Use Cases for mzQC

The following cases of usage have driven the development of the mzQC data model, and are used to define the scope of the format in version 0.1.0

mzQC is intended for:

1. reporting quality metrics calculated by QC tools (such as Quameter, rawMeat, …)
2. handover of quality reports of measured MS runs to researchers
3. storing / archiving QC metrics next to MS files / results in repositories / LIMS.
4. recording longitudinal QC with QC samples to monitor instrument health
5. QC of groups of measurements to monitor
   * method (acquisition) efficacy
   * Outlier detection
   * (analysis) precision
   * dataset “health”
6. explore and select datasets within repositories
7. visualisation input for reports / acquisition method debugging / improvements

In later versions specifying thresholds for QC metrics or flagging metrics / runs / sets of runs as “quality unfulfilled” is not modelled in the current version. Once this modelled, a mzQC file may be enriched in a second pass with thresholds and flags together with some rationale, why that “quality unfulfilled” flag was set.

# Notational Conventions

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 (3).

JSON data structures are built on two structures:

* a (unordered) collection of name/value pairs, which is various languages realized as an *object*.
* an (ordered) list of values, in most languages realized as an *array* or *list*.

An object begins with a *left brace* and ends with a *right brace*. Each name is followed by a *colon* and the name/value pairs are separated by a *comma.* An array begins with a *left bracket* and ends with a *right bracket*. Values are separated by a *comma*. They can be a *string* in double quotes, or a *number*, or **true** or **false** or **null**, or an *object* or an *array*. These structures can be nested.

A strings and values are like the respective C or Java structure. [json.org]

In respect to mass spectrometry-based experiments, one data acquisition resulting in one (vendor specific of MzML) peak list-file is referred to as a run.

# Relationship to Other Specifications

The mzQC format describes quality control information from mass spectrometry (MS)-based experiments and is therefore dependent on data in different formats. The specification of mzQC tries to minimise the replication of information contained in other PSI formats and uses concepts of and references to related specifications.

(2).

## The PSI Mass Spectrometry Controlled Vocabulary (MSCV)

The PSI-MS controlled vocabulary (6) is intended to provide terms for annotation of mzQC files. The CV has been generated with a 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:1002072 “p-value”) 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 with a “has\_units” key in the CV itself (relationship: has\_units: UO:0000221 ! dalton).

## The Quality Control Controlled Vocabulary (QCCV)

The PSI-QC controlled vocabulary is intended to provide terms for the definition of quality metrics and related supporting values. The CV has been generated initially with a collection of published and basic metrics. Further care went into the definition of the metrics and the involved values to help interpretation, use, and visualisation. The vocabulary builds on established terms and definitions from chemistry, physics, and biology ontologies in references and term relations.

As recommended by the PSI CV guidelines, psi-qccv.obo should be dynamically maintained via the [psidev-qc-dev@lists.sourceforge.net](mailto:psidev-ms-vocab@lists.sourceforge.net) mailing list that allows any user to request new terms, in agreement with the community involved.

Additionally, changes and new entries can be requested and discussed within a dedicated issue on the group’s development repository <https://github.com/HUPO-PSI/mzQC/issues> where a template is available to guide new entry definitions. Once a consensus is reached among the community the new terms are added within a few business days.

## MzML

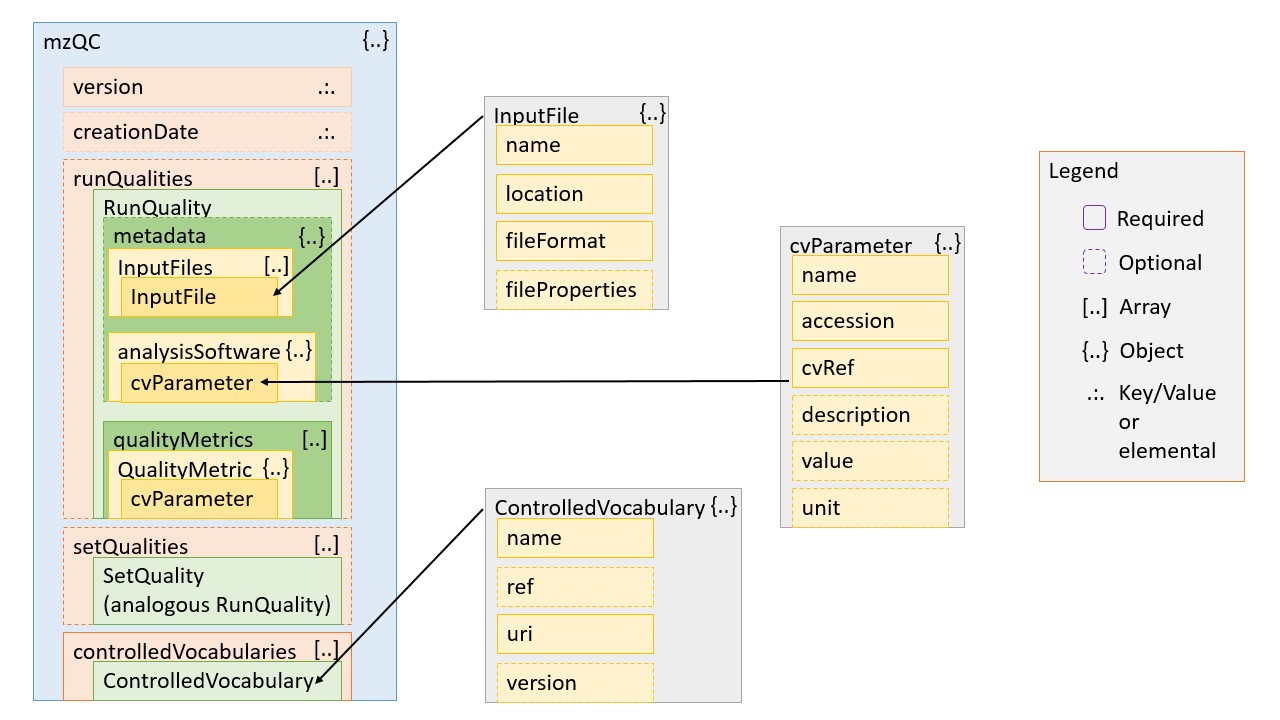
*mzML* (<http://www.psidev.info/mzml>) is the PSI standard for capturing mass spectra / peak lists resulting from MS in proteomics. It is RECOMMENDED that mzIdentML should be used in conjunction with mzML, although it is possible to use mzIdentML with other formats of mass spectra. As one of the main sources of data to compute quality metrics from, mzML has a strong relationship to mzQC. The order of reference however will predominantly be from mzQC to mzML, in particular the spectra and their nativeID in native spectrum identifier format as specified in MSCV (MS:1000767). The inverse is well possible, too, since metric entries in mzQC are CV term defined and mzML supports CV term annotation of it’s elements in many strategically important places of the schema. E.g. in case a quality based spectrum filter has been applied to an mzML and it needs to be further used or archived, the respective filter value or a reference to the quality metric can be stored in the <spectrum> or <spectrumList> elements <cvParam> in addition to the according <dataProcessing> entry.

## MzIdentML

*mzIdentML* (<http://www.psidev.info/mzidentml>) is the PSI standard for capturing of peptide and protein identification data (Jones, A. R., *et al.* 2012). As with mzML, the connection from mzQC to mzIdentML is strong. Much of the raw data for identification based quality metrics needs to be sourced from identification file formats, preferably mzIdentML, since as with mzML, references and values can be stored in mzIdentML’s <cvParam> elements of <SpectrumIdentificationItem>, <SpectrumIdentificationResult> and <SpectrumIdentificationList>.

In general, modifications SHOULD be sourced from Unimod (<http://www.unimod.org/obo/unimod.obo>) where possible. Units SHOULD be sourced from the Units of measurement ontology (<https://www.ebi.ac.uk/ols/ontologies/uo>).

# Format specification



## Schema sections

### runQuality

Parent element describing QC metrics for a single run.

**Type:** *[baseQuality](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "baseQuality|outline)*

**Example:**

{

"metadata": {

"inputFiles": [..],

"analysisSoftware": [..]

},

"qualityMetrics": [..]

}

### controlledVocabularies

Collection of controlled vocabulary elements.

**Type:** *array*

**Example:**

"controlledVocabularies": [..]

### runQualities

List of runQuality elements.

**Type:** *array*

**Example:**

"setQualities": [..]

### setQualities

List of setQuality elements.

**Type:** *array*

**Example:**

"setQualities": [..]

### qualityMetric

Parameter describing a QC metric. Value defined by the controlled vocabulary term. Optionally further constrained by the definition of the unit cvParameter.

**Type:** *[cvParameter](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "cvParameter|outline)*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| description | string | False | Description of the metric |
| cvRef | string | True | Reference to the controlled vocabulary that contains the parameter definition (pattern: "^[A-Z]+$") |
| unit | cvParameter | False | A controlled vocabulary element describing the unit of the metric value |
| accession | string | True | Accession number identifying the metric within its CV |
| value | Single value, list of values, tables of values, matrix of values | False | Value of the metric |
| name | string | True | Name of the controlled vocabulary element describing the metric |

**Example:**

{

"cvRef": "QC",

"accession": "QC:0000000",

"name": "Number of MS1 spectra",

"value": 5074

}

### setQuality

Parent element describing combined QC metrics for multiple runs.

**Type:** *[baseQuality](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "baseQuality|outline)*

**Example:**

{

"metadata": {

"inputFiles": [..],

"analysisSoftware": [..]

},

"qualityMetrics": [..]

}

### cvParameters

Optional cvParameter tags for the QC analysis.

**Type:** *array*

**Item types:** *[cvParameter](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "cvParameter|outline)*

**min/max:** *(1, -)*

**Example:**

"cvParapeters": [..]

### inputFiles

List of input files based on which the QC metrics were generated.

**Type:** *array*

**Item types:** *[inputFile](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "inputFile|outline)*

**min/max:** *(1, -)*

**Example:**

"inputFiles": [

{

"location": "/path/to/file/CPTAC\_CompRef\_00\_iTRAQ\_01\_2Feb12\_Cougar\_11-10-09.mzML",

"name": "CPTAC\_CompRef\_00\_iTRAQ\_01\_2Feb12\_Cougar\_11-10-09",

"fileFormat": {

"cvRef": "MS",

"accession": "MS:1000584",

"name": "mzML format"

},

"fileProperties": [..]

}

]

### analysisSoftware

Software tool(s) used to generate the QC metrics.

**Type:** *array*

**Item types:** *cvParameter & object*

**min/max:** *(1, -)*

***Object definition:***

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| version | string | True | Version number of the software tool. |
| url | string (uri) | True | Publicly accessible URI of the software tool (documentation). |

**Example:**

"analysisSoftware": [

{

"cvRef": "MS",

"accession": "MS:1000752",

"name": "TOPP software",

"version": "2.4",

"uri": "openms.de"

}

]

### metadata

Metadata describing the QC analysis.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| fileProvenance | string | False | Optional string describing/documenting the provenance of the file |
| [cvParameters](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "cvParameters|outline) | array | False | See there |
| [inputFiles](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "inputFiles|outline) | array | True | See there |
| [analysisSoftware](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "analysisSoftware|outline) | array | True | See there |

**Example:**

"metadata": {

"inputFiles": [..],

"analysisSoftware": [..]

}

### cvParameter

Base element containing a reference to a controlled vocabulary.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| description | string | False | Description of the parameter |
| cvRef | string | True | Reference to the controlled vocabulary that contains the parameter definition (pattern: "^[A-Z]+$") |
| unit | cvParameter | False | A controlled vocabulary element describing the unit of the parameter value |
| accession | string | True | Accession number identifying the parameter within its CV |
| value | any | False | Value of the parameter |
| name | string | True | Name of the controlled vocabulary element describing the parameter |

**Example:**

{

"cvRef": "MS",

"accession": "MS:1000569",

"name": "SHA-1",

"value": "76de62feccaaaadb608e89d897db57135e39ad87"

}

### controlledVocabulary

Element describing a controlled vocabulary.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| version | string | False | Version of the controlled vocabulary |
| ref | string | False | Reference string to denote the use of this specific controlled vocabulary by |
| name | string | True | Full name of the controlled vocabulary |
| uri | string | True | Publicly accessible URI of the controlled vocabulary |

**Example:**

{

"id": "MS",

"name": "Proteomics Standards Initiative Mass Spectrometry Ontology",

"version": "4.1.7",

"uri": "<https://github.com/HUPO-PSI/psi-ms-CV/blob/master/psi-ms.obo>"

}

### fileFormat

Type of input file.

**Type:** *[cvParameter](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "cvParameter|outline)*

**Example:**

"fileFormat": {

"cvRef": "MS",

"accession": "MS:1000584",

"name": "mzML format"

}

### fileProperties

Detailed properties of the input file.

**Type:** *array*

**Item types:** *cvParameter*

**min/max:** *(1, -)*

**Recommendations:** The MS:1000747 term for completion time of a run SHOULD be included. The MS:1000031 term for completion time of a run SHOULD be included. A checksum term for the checksum of a run’s original peak list file SHOULD be included.

**Example:**

"fileProperties": [

{

"cvRef": "MS",

"accession": "MS:1000747",

"name": "completion time",

"value": "2012-02-03 11:00:41"

}

]

### inputFile

File based on which QC metrics were generated.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| [fileFormat](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "fileFormat|outline) | object | True | See there |
| [fileProperties](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "fileProperties|outline) | array | False | See there |
| location | string | True | Unique file location, preferably publicly accessible |
| name | string | True | Base file name, should be unique across the mzQC file |

**Example:**

{

"location": "/path/to/file/CPTAC\_CompRef\_00\_iTRAQ\_01\_2Feb12\_Cougar\_11-10-09.mzML",

"name": "CPTAC\_CompRef\_00\_iTRAQ\_01\_2Feb12\_Cougar\_11-10-09",

"fileFormat": {

"cvRef": "MS",

"accession": "MS:1000584",

"name": "mzML format"

},

"fileProperties": [

{

"cvRef": "MS",

"accession": "MS:1000747",

"name": "completion time",

"value": "2012-02-03 11:00:41"

},

{

"cvRef": "MS",

"accession": "MS:1000569",

"name": "SHA-1",

"value": "76de62feccaaaadb608e89d897db57135e39ad87"

},

{

"cvRef": "MS",

"accession": "MS:1000031",

"name": "instrument model",

"value": "LTQ Orbitrap Velos"

}

]

}

### qualityMetrics

The collection of quality metrics for a particular run / group of runs

**Type:** *array*

**Item types:** *[qualityMetric](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "qualityMetric|outline)*

**min/max:** *(1, -)*

**Example:**

"qualityMetrics": [..]

### baseQuality

Base element describing a runQuality or setQuality.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| [qualityMetrics](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "qualityMetrics|outline) | array | True | See there |
| [metadata](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "metadata|outline) | object | True | See there |

**Example:** see runQuality or setQuality.

### mzQC

Basis of a mzQC file.

**Type:** *object*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Required | Description |
| creationDate | string | True | Creation date of the file |
| version | string | True | Version of the format |
| [runQualities](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "qualityMetrics|outline) | array | True | See there |
| [setQualities](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "metadata|outline) | array | True | See there |
| [controlledVocabularies](https://ukc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en-GB&rs=en-US&wopisrc=https%3A%2F%2Femblebi-my.sharepoint.com%2Fpersonal%2Fwalzer_ebi_ac_uk%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F6b7175ad6e8b46f3b79c4bea20aea29c&wdenableroaming=1&mscc=1&wdodb=1&hid=F0FE189F-50C6-2000-25DD-4C359834C6C9&wdorigin=Other&jsapi=1&newsession=1&corrid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&usid=36982d44-39cf-4e63-b233-e42c9cb9b0e9&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush" \l "metadata|outline) | array | True | See there |

**Example:**

{

"mzQC": {

"creationDate": "2019-10-29T14:40:17",

"version": "0\_0\_11",

"runQualities": [..],

"setQualities": [..],

"controlledVocabularies": [..]

}

}

## CvParameter Values for metrics

The values of metrics are defined by the respective the controlled vocabulary term. Their type can be one of four different specifications.

**Type:** *string/number/boolean/array/object*

|  |  |  |
| --- | --- | --- |
| Definition | Type | example |
| Single value | string/number/boolean | 1.0 |
| List of values | array of values | [2,3,5,7,11] |
| Table of values | Object; each column a key-value pair of *column name: array of values*  All columns need to be of equal length | {score: [0,1,1,2,3,5], intensity: [0,1,3,6,10,15]} |
| Matrix of values | Array of array of values; each sub-array contains a row of the matrix | [[1,2,3], [4,5,6], [7,8,9]] |

# Conclusions

This document contains the specifications for using the mzQC format to represent quality control metrics in the context of a mass spectrometry-based experimentation. This specification constitutes a proposal for a standard from the Proteomics Standards Initiative. These artefacts are currently undergoing the PSI document process, which will result in a standard officially sanctioned by PSI.

# Authors

Mathias Walzer

European Bioinformatics Institute (EMBL-EBI)

Hinxton, United Kingdom

[walzer@ebi.ac.uk](mailto:Walzer@ebi.ac.uk)

Correspondence – Mathias Walzer ([walzer@ebi.ac.uk)](mailto:tobias@ebi.ac.uk)).

# Contributors

In addition to the authors, the following people contributed to the model development, gave feedback or tested:

Martin Eisenacher

# References

1. Tyner C.; Barber, G.P.; Casper J.; Clawson, H.; Diekhans, M.; Eisenhart, C.; Fischer, C.M.; Gibson, D.; Gonzalez, J.N.; Guruvadoo, L.; Haeussler, M.; Heitner, S.; Hinrichs, A.S.; Karolchik, D.; Lee, B.T.; Lee, C.M.; Nejad, P.; Raney, B.J.; Rosenbloom, K.R.; Speir, M.L.; Villarreal, C.; Vivian, J.; Zweig, A.S.; Haussler, D.; Kuhn, R.M.; Kent, W.J. The UCSC GenomeBrowser database: 2017 update. *Nucleic Acids Res.* **2017**, 45(D1), D626-34.

2. Kent, W. J.; Zweig, A. S.; Barber, G.; Hinrichs, A. S.; Karolchik, D., BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **2010,** 26, (17), 2204-7.

3. Bradner, S; Key words for use in RFCs to Indicate Requirement Levels. In Internet Engineering Task Force. RFC 2119: **1997**.

4. Jones, A. R.; Eisenacher, M.; Mayer, G.; Kohlbacher, O.; Siepen, J.; Hubbard, S. J.; Selley, J. N.; Searle, B. C.; Shofstahl, J.; Seymour, S. L.; Julian, R.; Binz, P. A.; Deutsch, E. W.; Hermjakob, H.; Reisinger, F.; Griss, J.; Vizcaino, J. A.; Chambers, M.; Pizarro, A.; Creasy, D., The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol Cell Proteomics* **2012,** 11, (7), M111 014381.

5. Griss, J.; Jones, A. R.; Sachsenberg, T.; Walzer, M.; Gatto, L.; Hartler, J.; Thallinger, G. G.; Salek, R. M.; Steinbeck, C.; Neuhauser, N.; Cox, J.; Neumann, S.; Fan, J.; Reisinger, F.; Xu, Q. W.; Del Toro, N.; Perez-Riverol, Y.; Ghali, F.; Bandeira, N.; Xenarios, I.; Kohlbacher, O.; Vizcaino, J. A.; Hermjakob, H., The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Mol Cell Proteomics* **2014,** 13, (10), 2765-75.

6. Mayer, G.; Montecchi-Palazzi, L.; Ovelleiro, D.; Jones, A. R.; Binz, P.-A.; Deutsch, E. W.; Chambers, M.; Kallhardt, M.; Levander, F.; Shofstahl, J.; Orchard, S.; Antonio Vizcaíno, J.; Hermjakob, H.; Stephan, C.; Meyer, H. E.; Eisenacher, M., The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary. *Database* **2013,** 2013.

7. Ghali, F.; Krishna, R.; Perkins, S.; Collins, A.; Xia, D.; Wastling, J.; Jones, A. R., ProteoAnnotator--open source proteogenomics annotation software supporting PSI standards. *Proteomics* **2014,** 14, (23-24), 2731-41.

# Appendix I: qcML to mzQC conversion

Conversion from a qcML file to the mzQC format is possible.

The straight forward case where old metrics are available in the PSI QC working group’s controlled vocabulary,

(2)

## Incompatible metrics

# 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).

# TradeMark Section

Microsoft Excel®

# Copyright Notice

Copyright (C) Proteomics Standards Initiative (2017-20). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the PSI or other organizations, except as needed for the purpose of developing Proteomics Recommendations in which case the procedures for copyrights defined in the PSI Document process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the PSI or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE PROTEOMICS STANDARDS INITIATIVE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE."