Justus Liebig University, Giessen
Institute of Inorganic and Analytical Chemistry

Thorsten Schramm
Alfons Hester
Andreas Römpp

August 2009

imzML: Imaging Mass Spectrometry Markup Language

Status of this document

This document presents the version 1.1.0 RC1 of the specification for the imzML (Imaging Mass Spectrometry Markup Language) data format developed in the course of the EU funded project COMPUTIS. Distribution is unlimited.

Version of this document

The current version of this document is: Version 1.1.0 RC1; August 31, 2009.

Contents

1. Introduction

1.1 Background

Imaging mass spectrometry is the method of scanning a sample of interest and generating an ‘image’ of the intensity distribution of a specific analyte/ion. This method results in a large number of spectra which are typically acquired with identical measurement parameters. The application of imaging MS is rapidly growing with a constantly increasing number of different instrumental systems and software tools.

COMPUTIS is an EU funded project that develops new and improved technologies for molecular imaging mass spectrometry [1]. An important task within this project is the comparison of images generated by diverse types of mass spectrometers. Both the DICOM standard for in-vivo imaging data [2] and the mzML standard by HUPO-PSI [3, 4] are not able to completely represent an imaging MS experiment. Therefore a standardized data format was developed to simplify the exchange of imaging MS data between different instrument and data analysis software.
Several previous imaging MS data formats utilize two separate files: a small (ini or XML) file for the metadata and a larger (binary) file for the mass spectral data (biomap, DCE, udp). We decided to keep this structure in order to ensure flexible and fast handling of the imaging MS data. We also decided that the metadata file should be based on the mass spectrometry standard mzML developed by HUPO-PSI (http://psidev.info/index.hp?q=node/257). The MS data is stored in one of two binary formats in order to ensure the most efficient storage of these large data sets (section 3.2). A new controlled vocabulary was compiled for imzML to include parameters that are specific for imaging experiments. These parameters are stored in the imagingMS.obo file.

Why not use mzML?

First of all mzML was not available at the beginning of the COMPUTIS project (2006). At later stages we evaluated mzML with respect to imaging MS data. The main concern was the file size of converted data sets. Storing the MS data in a separate binary file is crucial for handling the very large imaging MS data sets. After it became clear that our main requirement (a separate binary file) was not possible in mzML, we decided to continue with our own format. However, we decided to store our metadata in the mzML format in order to be able to easily convert between the two formats. During the last years we stayed in contact with HUPO-PSI at various occasions. The result of these discussions was that we call our format imzML (for imaging mzML) and that it will exist in parallel to mzML for specific use cases (e.g. large data sets). The structure of the XML metadata file will remain compatible with mzML 1.1. The imaging specific cv parameters will be kept in a separate imagingMS obo file. A number of entries from this file (which were of general importance) have already been included in the psi-ms obo file.

1.2 Previous formats

imzML was developed on the basis of mzML and previous imaging MS data formats such as biomap and internal formats developed at JLU and AMOLF

imzML is based on mzML () and
As the XML structure of imzML is basically identical with mzML (with the exception of empty binary tags and additional cv params), this document is limited to the changes with respect to mzML. Further information on mzML can be found in the mzML version 1.1.0 documentation (http://psidev.info/index.php?q=node/257). The specifications of the binary MS data file are described in section 3.2.

1.3 Design Philosophy

The fundamental goal of the imzML developers was to design a data format for the exchange of imaging MS data. At the same time the format should be easily interchangeable with mzML.
The main goals can be summarized as

1. Ensure complete description of imaging MS experiments
2. Minimize file size
3. Ensure fast and flexible data handling
4. keep the (XML part of) imzML as close as possible to mzML 1.1

2. Implementation of the format

2.1 Concepts and terminology

This document assumes familiarity with one data modelling notation, namely XML Schema (www.w3.org/XML/Schema). Models are described using XML schema.

The keywords “MUST,” “MUST NOT,” “REQUIRED,” “SHALL,” “SHALL NOT,” “SHOULD,” “SHOULD NOT,” “RECOMMENDED,” “MAY,” and “OPTIONAL” are to be interpreted as described in RFC-2119 (Bradner 1997).

2.2 Relationship to other specifications

Since imzML is based on mzML the relations to the format and models of mzML is connected with are also relevant for imzML. See mzML documentation 1.1.0 for further details.

2.2.1 Differences between mzML 1.1.0 and imzML 1.1.0

The imzML files created according to this documentation will pass the regular mzML validators i.e. the OpenMS mzML validator (http://www-bs2.informatik.uni-tuebingen.de/services/OpenMS/mzML/) or the

There were some further modifications necessary to validate against mzML:

- The <binary> is now empty and holds no information
- The attributes encodedLength (defaultEncodedLength) and arrayLength (defaultArrayLength) are always 0. The real sizes are given via a <cvParam> within the <binaryDataArray>
- The information about the maximum image size and scanning procedure are stored in the newly renamed <scanSettings> section.
- The cvParams describing the position of the spectrum in the image are saved in the in the <scan> section of the <spectrum>.
- cvParams describing the instrumentation source are saved in the <source> section of the <instrumentConfigurationList>

ProDaC on-line validator (http://eddie.thep.lu.se/prodac_validator/validator.pl). The binary tag has to be empty. The additional imaging parameters of the added controlled vocabulary will create warning messages, but no errors.

2.3 The imaging mass spectrometry controlled vocabulary (cv)

The base for the imaging MS controlled vocabulary is the psi-ms.obo. The new controlled vocabulary contains a collection of terms which are necessary to distinctly describe an imaging experiment so that it can be interpretated unequivocaly and reproduced exactly. The first branch terms (the direct children of the root term) can be separated into two categories: one for the dealing with the external files and one organising the imaging parameters.

The controlled vocabulary is maintained by the Institute of Inorganic and Analytical Chemistry at the Justus Liebig University, Giessen. If somebody wants to change, add or remove a term contact Andreas Römpp via email (andreas.roempp@uni-giessen.de). If a change is confirmed, the new controlled vocabulary will be available for download a few days later. The name of the file will not change but the number of the version. For this reason an internal version number with three decimals (x.y.z) should be increased (taken from the mzML description):

- x should be increased when a first level term are renamed added deleted or rearranged in the structure. Such rearrangement is suppose to 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 are required in order to use the ImagingMS.obo:

- Unit Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=unit)
- PSI-MS Ontology (http://psidev.cvs.sourceforge.net/*checkout*/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo) - see mzML documentation 1.1.0 for further details.

2.4 Orientation of images

Every spectrum contains information about the x and y axis position. These parameters are part of the controlled vocabulary. It is not necessary to know, in which way the image was generated (although this information will be stored as well in the imzML file), because the position of every spectrum is distinctly given. The beginning of both x and y axis is in the top left corner. This position is defined as 1/1 (see figure 1). This way every application should generate images of identical orientation.

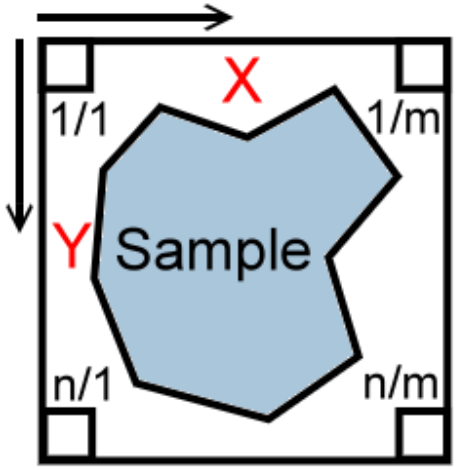

Figure 1: Orientation of x and y axis. The x-axis is the horizontal one, the y-axis is vertically orientated. The position 1/1 is in the top left corner.

2.5 Other supporting materials

To completely understand this documentation it is necessary to take a look at the mzML documentation, the controlled vocabularies of PSI-MS and ImagingMS and the example file of imzML.

All information, files and programs for **mzML** are available at:
http://psidev.info/index.php?q=node/257

All information and files of **imzML** are available at the http://www.maldi-msi.org/.

They are:

Filename	Description
imzML1.1.0.xsd	Main imzML XML schema definition file
imagingMS.obo	controlled vocabulary of imaging mass spectrometry in OBO format
imzML1.1-mapping.txt	XML-encoded rules for where certain cvParams MAY/MUST

	appear in the document.
Example_Processed.zip	imzML and ibd file saved in the processed format
Example_Continuous.zip	imzML and ibd file saved in the continuous format
Example_Images.zip	A collection of 8 images showing 8 different m/z ratios of the example imzML files (continuous and processed)

Additional material will be referenced with hyperlinks at the same URL.

3. File structure

The imzML file holds the metadata of an MS image which is described by the mzML based XML structure and the extended controlled vocabulary. The imaging binary data file (*.ibd) contains the mass spectral data. The connection between the two files is made via links in the XML file which hold the offsets of the mass spectral data in the binary file. It is important to keep in mind that the information of both file is only valid if no file is missing. Therefore the user should be very careful when copying or moving those files; inaccurate file handling can result in data loss. It is recommend to keep the files together and use the same names for the imzML part and the ibd part.

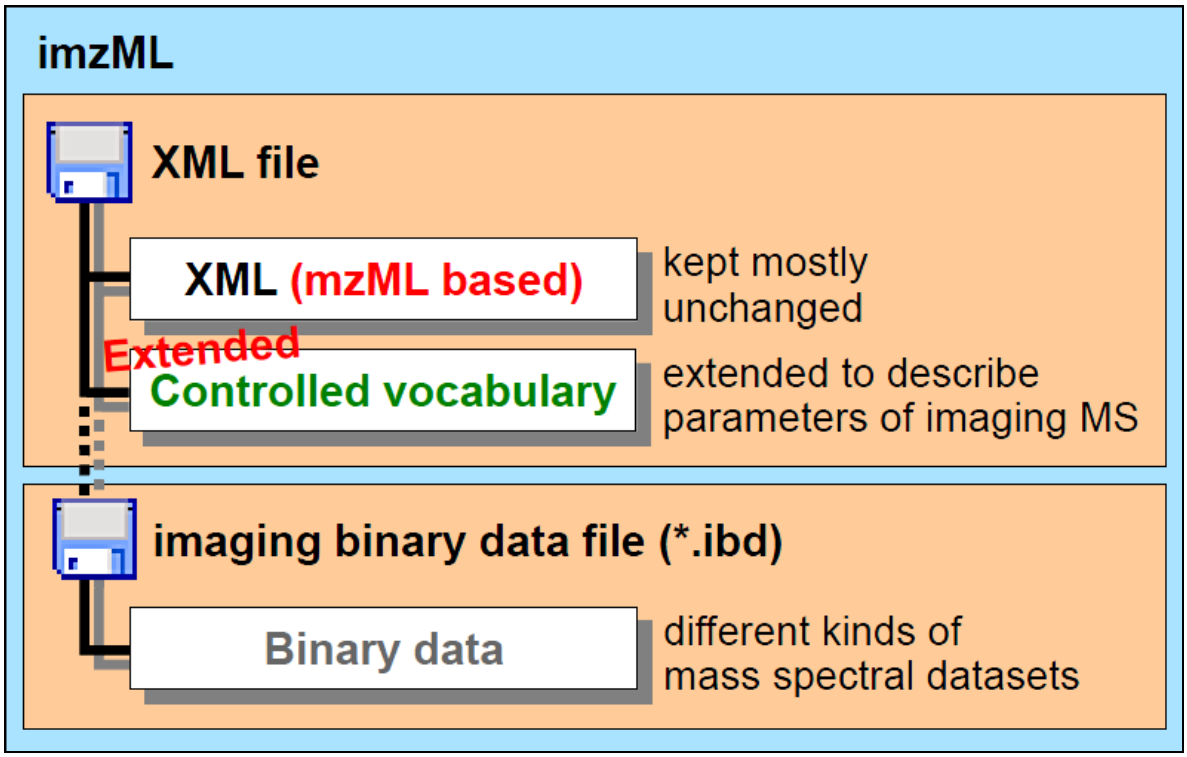

Figure 2: General structure of imzML.

Both files carry the same universally unique identifier (UUID) (http://tools.ietf.org/html/rfc4122). This can help to reunite lost files. The UUID is defined in the controlled vocabulary (imagingMS.obo) and this parameter is located in the imzML file in the <fileContent> tag. The first 16 byte of the imaging binary data file contain the same UUID. Comparing both UUIDs allows to find out if the two files build a pair of corresponding files.

3.1 The XML file

The XML model of imzML is the same as for mzML. The general XML structure of mzML is kept unchanged. See the mzML version 1.1.0 documentation (section 3) for further details.
There is only one XML element which differs slightly from the original mzML: The function of the <binary/> tag was altered in such a way that it now contains no base64 encoded binary data. It stays empty, which is compatible to mzML 1.1.0. This results in constant values for “encoded length” and “array length” of zero in the parent tags such as <spectrum/> and <binaryDataArray/>. Most of the changes in the XML part are related to cvParam mapping rules for the newly introduced parameters of the imaging controlled vocabulary (see section 4).

3.2 The binary file

The first 16 bytes of the binary file are reserved for an Universally Unique Identifier (UUID). It is also saved in the imzML file so that a correct assignment of ibd and imzML file is possible even if the names of both files are different.
Two basic formats of the binary file exist: continuous and processed. Continuous type means that each spectrum of an image has the same m/z values. As a result the m/z array is only saved once directly behind the UUID of the file and the intensity arrays of the spectra are following. At the processed type every spectrum has its own m/z array. So it is necessary to save both – the m/z array and the corresponding intensity array – per spectrum.

Integer values may be stored as byte (1 byte), word (2 bytes), longword (4 bytes), long (8 bytes). Floating point values may be stored as single or double (cf. IEEE 754). The byte order is **always** little endian (intel style).

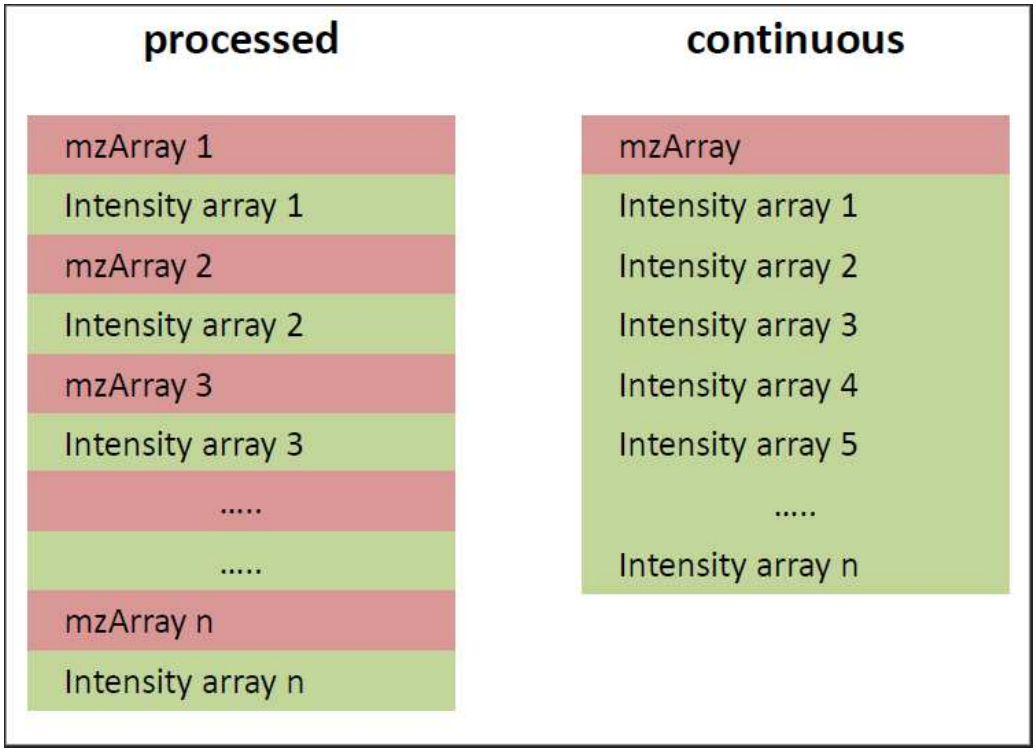

Figure 3: The two possibilities to store mass spectral data in the imaging binary data file: continuous and processed

4. Changed mzML elements

Since the mass spectrometric data is saved in an external binary file there will be **no** data saved in the XML file. Therefore the <binary> tag will **never** contain any data! All other attributes of the XML tag, which give information about the characteristics of the data array have to be set to ZERO. Otherwise the file will not be validated correctly.

Only those elements are explained in detail where changes happened in comparison to mzML. The changes are marked by *red and cursive* writing style.

4.1 Element <fileContent>

Definition: This summarizes the different types of spectra that can be expected in the file. This is expected to aid processing software in skipping files that do not contain appropriate spectrum types for it. It should also describe the nativeID format used in the file by referring to an appropriate CV term.

Type: dx:ParamGroupType

Attributes: none

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead

Example Context:

```
<fileContent>
  <cvParam cvRef="MS" accession="MS:1000580" name="MSn spectrum" value=""/>
  <userParam name="ProteoWizard" value="Thermo RAW data converted to mzML, with additional
MIAPE parameters added for illustration"/>
</fileContent>
```

cvParam Mapping Rules:

Path mzML/fileDescription/fileContent

```
MUST supply a *child* term of MS:1000524 (data file content) one or more times
  e.g.: MS:1000235 (total ion current chromatogram)
  e.g.: MS:1000322 (charge inversion mass spectrum)
  e.g.: MS:1000325 (constant neutral gain spectrum)
  e.g.: MS:1000326 (constant neutral loss spectrum)
  e.g.: MS:1000328 (e/2 mass spectrum)
  e.g.: MS:1000341 (precursor ion spectrum)
  et al.
MAY supply a *child* term of MS:1000525 (spectrum representation) only once
  e.g.: MS:1000127 (centroid spectrum)
  e.g.: MS:1000128 (profile spectrum)
MUST supply a *child* term of IMS:1000003 (ibd binary type) only once
  e.g.: IMS:1000030 (continuous)
  e.g.: IMS:1000031 (processed)
MUST supply a *child* term IMS:1000009 (ibd checksum) only once
  e.g.: IMS:1000090 (ibd MD5)
  e.g.: IMS:1000091 (ibd SHA-1)
MUST supply a *child* term IMS:1000008 (ibd identification) only once
  e.g.: IMS:1000080 (universally unique identifier)
MAY supply a *child* term IMS:1000007 (ibd file) only once
  e.g.: IMS:1000070 (external binary uri)
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000580" name="MSn spectrum" value=""/>
<cvParam cvRef="MS" accession="MS:1000127" name="centroid spectrum" value=""/>
<cvParam cvRef="MS" accession="MS:1000326" name="constant neutral loss spectrum"/>
```

4.2 Element <spectrum>

Definition: The structure that captures the generation of a peak list (including the underlying acquisitions). Also describes some of the parameters for the mass spectrometer for a given acquisition (or list of acquisitions).

Type: dx:SpectrumType

Attributes:

Attribute Name	Data Type	Use	Definition
dataProcessingRef	xs:IDREF	optional	This attribute can optionally reference the 'id' of the appropriate dataProcessing.
defaultArrayLength	xs:int	required	Default length of binary data arrays contained in this element. It has to be always set to ZERO!
id	xs:string (pattern: \S+=\S+(\S+=\S+)*)	required	The native identifier for a spectrum. For unmerged native spectra or spectra from older open file formats, the format of the identifier is defined in the PSI-MS CV and referred to in the mzML header. External documents may use this identifier together with the mzML filename or accession to reference a particular spectrum.
index	xs:nonNegativeInteger	required	The zero-based, consecutive index of the spectrum in the SpectrumList.
sourceFileRef	xs:IDREF	optional	This attribute can optionally reference the 'id' of the appropriate sourceFile.
spotID	xs:string	optional	The identifier for the spot from which this spectrum was derived, if a MALDI or similar run.

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead
scanList	0	1	List and descriptions of scans.

precursorList	0	1	List and descriptions of precursor isolations to the spectrum currently being described, ordered.
productList	0	1	List and descriptions of product isolations to the spectrum currently being described, ordered.
binaryDataArrayList	0	1	List of binary data arrays.

Example Context:

```
<spectrum index="3" id="scan=22" spotID="A1,42x42,4242x4242" defaultArrayLength="15">
  <referenceableParamGroupRef ref="CommonMS1SpectrumParams"/>
  <cvParam cvRef="MS" accession="MS:1000511" name="ms level" value="1"/>
  <cvParam cvRef="MS" accession="MS:1000127" name="centroid spectrum" value=""/>
  <cvParam cvRef="MS" accession="MS:1000528" name="lowest observed m/z"
value="142.38999999999999" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  <cvParam cvRef="MS" accession="MS:1000527" name="highest observed m/z"
value="942.55999999999995" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  <cvParam cvRef="MS" accession="MS:1000504" name="base peak m/z" value="422.42000000000002"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  ...
</spectrum>
```

cvParam Mapping Rules:

Path mzML/run/spectrumList/spectrum

```
MAY supply a *child* term of MS:1000465 (scan polarity) only once
  e.g.: MS:1000129 (negative scan)
  e.g.: MS:1000130 (positive scan)
MUST supply a *child* term of MS:1000559 (spectrum type) only once
  e.g.: MS:1000322 (charge inversion mass spectrum)
  e.g.: MS:1000325 (constant neutral gain spectrum)
  e.g.: MS:1000326 (constant neutral loss spectrum)
  e.g.: MS:1000328 (e/2 mass spectrum)
  e.g.: MS:1000341 (precursor ion spectrum)
  e.g.: MS:1000581 (CRM spectrum)
  e.g.: MS:1000582 (SIM spectrum)
  e.g.: MS:1000583 (SRM spectrum)
  e.g.: MS:1000789 (enhanced multiply charged spectrum)
  e.g.: MS:1000790 (time-delayed fragmentation spectrum)
  et al.
MUST supply term MS:1000525 (spectrum representation) or any of its children only once
  e.g.: MS:1000127 (centroid spectrum)
  e.g.: MS:1000128 (profile spectrum)
MAY supply a *child* term of MS:1000499 (spectrum attribute) one or more times
  e.g.: MS:1000285 (total ion current)
  e.g.: MS:1000497 (zoom scan)
  e.g.: MS:1000504 (base peak m/z)
  e.g.: MS:1000505 (base peak intensity)
  e.g.: MS:1000511 (ms level)
  e.g.: MS:1000527 (highest observed m/z)
  e.g.: MS:1000528 (lowest observed m/z)
  e.g.: MS:1000618 (highest observed wavelength)
  e.g.: MS:1000619 (lowest observed wavelength)
  e.g.: MS:1000796 (spectrum title)
  et al.
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000580" name="MSn spectrum"/>
<cvParam cvRef="MS" accession="MS:1000511" name="ms level" value="2"/>
<cvParam cvRef="MS" accession="MS:1000127" name="centroid spectrum"/>
<cvParam cvRef="MS" accession="MS:1000130" name="positive scan"/>
<cvParam cvRef="MS" accession="MS:1000285" name="total ion current" value="1.0289517E7"/>
<cvParam cvRef="MS" accession="MS:1000128" name="profile spectrum" value=""/>
<cvParam cvRef="MS" accession="MS:1000504" name="base peak m/z" value="810.415283203125"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
<cvParam cvRef="MS" accession="MS:1000505" name="base peak intensity" value="1471973.875"
unitCvRef="MS" unitAccession="MS:1000131" unitName="number of counts"/>
<cvParam cvRef="MS" accession="MS:1000528" name="lowest observed m/z"
value="200.00018816645022" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
<cvParam cvRef="MS" accession="MS:1000527" name="highest observed m/z"
value="2000.0099466203771" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
<cvParam cvRef="MS" accession="MS:1000326" name="constant neutral loss spectrum"/>
```

Notes and Constraints:

id's MUST be unique within a file as constrained by a primary key. The format MUST follow the native ID guidelines for mzML If a scan yields no peaks, it should still be reported, but with a defaultArrayLength of 0 and no <binaryDataArrayList> element.

4.3 Element <scanSettings>

Definition: Description of the acquisition settings of the instrument prior to the start of the run.

Type: dx:ScanSettingsType

Attributes:

Attribute Name	Data Type	Use	Definition
id	xs:ID	required	A unique identifier for this acquisition setting.

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead
sourceFileRefList	0	1	List with the source files containing the acquisition settings.
targetList	0	1	Target list (or 'inclusion list') configured prior to the run.

Example Context:

```
<scanSettings id="as1">
  <sourceFileRefList count="1">
    <sourceFileRef ref="sf_parameters"/>
  </sourceFileRefList>
  <targetList count="2">
    <target>
      <cvParam cvRef="MS" accession="MS:1000744" name="selected ion m/z" value="1000"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  ...
</scanSettings>
```

cvParam Mapping Rules:

Path mzML/scanSettingsList/scanSettings

```
MUST supply a *child* term of IMS:1000049 (line scan direction) only once
  e.g.: IMS:1000492 (line scan bottom up)
  e.g.: IMS:1000491 (line scan left right)
  e.g.: IMS:1000490 (line scan right left)
  e.g.: IMS:1000493 (line scan top down)
MUST supply a *child* term of IMS:1000040 (scan direction) only once
  e.g.: IMS:1000400 (bottom up)
  e.g.: IMS:1000402 (left right)
  e.g.: IMS:1000403 (right left)
  e.g.: IMS:1000401 (top down)
MUST supply a *child* term of IMS:1000041 (scan pattern) only once
  e.g.: IMS:1000410 (meandering)
  e.g.: IMS:1000411 (one way)
  e.g.: IMS:1000412 (random access)
MUST supply a *child* term of IMS:1000048 (scan type) only once
  e.g.: IMS:1000480 (horizontal line scan)
  e.g.: IMS:1000481 (vertical line scan)
SHOULD supply a *child* term of IMS:1000040 (image) one or more times
  e.g.: IMS:1000044 (max dimension x)
  e.g.: IMS:1000045 (max dimension y)
  e.g.: IMS:1000046 (pixel size)
```

Example cvParams:

```
<cvParam cvRef="IMS" accession="MS:1000491" name="line scan left right" value=""/>
<cvParam cvRef="IMS" accession="MS:1000402" name="top down" value=""/>
<cvParam cvRef="IMS" accession="MS:1000411" name="one way" value=""/>
<cvParam cvRef="IMS" accession="MS:1000480" name="horizontal line scan" value=""/>
```

4.4 Element <source>

Definition: A source component.

Type: dx:SourceComponentType

Attributes:

Attribute Name	Data Type	Use	Definition
order	xs:int	required	This attribute MUST be used to indicate the order in which the components are encountered from source to detector (e.g., in a Q-TOF, the quadrupole would have the lower order number, and the TOF the higher number of the two).

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead

Example Context:

```
<source order="1">
  <cvParam cvRef="MS" accession="MS:1000073" name="electrospray ionization" value=""/>
  <cvParam cvRef="MS" accession="MS:1000057" name="electrospray inlet" value=""/>
  <cvParam cvRef="MS" accession="MS:1000486" name="source potential" value="4.20"
unitCvRef="UO" unitAccession="UO:0000218" unitName="volt"/>
</source>
```

cvParam Mapping Rules:

Path mzML/instrumentConfigurationList/instrumentConfiguration/componentList/source

```
MAY supply a *child* term of MS:1000482 (source attribute) one or more times
  e.g.: MS:1000392 (ionization efficiency)
  e.g.: MS:1000486 (source potential)
  e.g.: MS:1000843 (wavelength)
  e.g.: MS:1000844 (focus diameter x)
  e.g.: MS:1000845 (focus diameter y)
  e.g.: MS:1000846 (pulse energy)
  e.g.: MS:1000847 (pulse duration)
  e.g.: MS:1000848 (attenuation)
  e.g.: MS:1000849 (impact angle)
  e.g.: MS:1000850 (gas laser)
  et al.
MUST supply term MS:1000008 (ionization type) or any of its children only once
  e.g.: MS:1000070 (atmospheric pressure chemical ionization)
  e.g.: MS:1000071 (chemical ionization)
  e.g.: MS:1000074 (fast atom bombardment ionization)
  e.g.: MS:1000075 (matrix-assisted laser desorption ionization)
  e.g.: MS:1000227 (multiphoton ionization)
  e.g.: MS:1000239 (atmospheric pressure matrix-assisted laser desorption ionization)
  e.g.: MS:1000255 (flowing afterglow)
  e.g.: MS:1000257 (field desorption)
  e.g.: MS:1000258 (field ionization)
  e.g.: MS:1000259 (glow discharge ionization)
  et al.
MAY supply a *child* term of MS:1000007 (inlet type) only once
  e.g.: MS:1000055 (continuous flow fast atom bombardment)
  e.g.: MS:1000056 (direct inlet)
  e.g.: MS:1000058 (flow injection analysis)
  e.g.: MS:1000059 (inductively coupled plasma)
  e.g.: MS:1000060 (infusion)
  e.g.: MS:1000061 (jet separator)
  e.g.: MS:1000062 (membrane separator)
  e.g.: MS:1000063 (moving belt)
  e.g.: MS:1000064 (moving wire)
  e.g.: MS:1000065 (open split)
  et al.
MAY supply a *child* term of IMS:1000002 (sample stage) one or more times
  e.g.: IMS:1000200 (position accuracy)
  e.g.: IMS:1000201 (step size)
  e.g.: IMS:1000202 (target material)
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000398" name="nanoelectrospray" value=""/>
<cvParam cvRef="MS" accession="MS:1000073" name="electrospray ionization" value=""/>
<cvParam cvRef="MS" accession="MS:1000057" name="electrospray inlet" value=""/>
<cvParam cvRef="MS" accession="MS:1000486" name="source potential" value="4.20" unitCvRef="UO"
unitAccession="UO:0000218" unitName="volt"/>
```

4.5 Element <scan>

Definition: Scan or acquisition from original raw file used to create this peak list, as specified in sourceFile.

Type: dx:ScanType

Attributes:

Attribute Name	Data Type	Use	Definition
externalSpectrumID	xs:string	optional	For scans that are external to this document, this string MUST correspond to the 'id' attribute of a spectrum in the external document indicated by 'sourceFileRef'.
instrumentConfigurationRef	xs:IDREF	optional	This attribute can optionally reference the 'id' attribute of the appropriate instrument configuration.
sourceFileRef	xs:IDREF	optional	If this attribute is set, it MUST reference the 'id' attribute of a sourceFile representing the external document containing the spectrum referred to by 'externalSpectrumID'.
spectrumRef	xs:string	optional	For scans that are local to this document, this attribute can be used to reference the 'id' attribute of the spectrum corresponding to the scan.

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead
scanWindowList	0	1	Container for a list of scan windows.

Example Context:

```
<scan instrumentConfigurationRef="LCQDeca">
  <cvParam cvRef="MS" accession="MS:1000016" name="scan start time"
value="42.049999999999997" unitCvRef="UO" unitAccession="UO:0000010" unitName="second"/>
  <cvParam cvRef="MS" accession="MS:1000512" name="filter string" value="+ c MALDI Full ms
[100.00-1000.00]"/>
  <scanWindowList count="1">
    <scanWindow>
      <cvParam cvRef="MS" accession="MS:1000501" name="scan window lower limit" value="100"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
      <cvParam cvRef="MS" accession="MS:1000500" name="scan window upper limit" value="1000"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  ...
</scan>
```

cvParam Mapping Rules:

Path mzML/run/spectrumList/spectrum/scanList/scan

```
MAY supply a *child* term of MS:1000503 (scan attribute) one or more times
  e.g.: MS:1000011 (mass resolution)
  e.g.: MS:1000015 (scan rate)
  e.g.: MS:1000016 (scan start time)
  e.g.: MS:1000502 (dwell time)
  e.g.: MS:1000512 (filter string)
  e.g.: MS:1000616 (preset scan configuration)
  e.g.: MS:1000800 (mass resolving power)
  e.g.: MS:1000803 (analyzer scan offset)
  e.g.: MS:1000826 (elution time)
  e.g.: MS:1000880 (interchannel delay)
MAY supply a *child* term of MS:1000018 (scan direction) only once
  e.g.: MS:1000092 (decreasing m/z scan)
  e.g.: MS:1000093 (increasing m/z scan)
MAY supply a *child* term of MS:1000019 (scan law) only once
  e.g.: MS:1000094 (exponential)
  e.g.: MS:1000095 (linear)
  e.g.: MS:1000096 (quadratic)
MUST supply term of IMS:1000050 (position x) only once
MUST supply term of IMS:1000051 (position y) only once
MAY supply term of IMS:1000052 (position z) only once
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000016" name="scan start time" value="5.8905000000000003"
unitCvRef="UO" unitAccession="UO:0000031" unitName="minute"/>
<cvParam cvRef="MS" accession="MS:1000512" name="filter string" value="+ c NSI Full ms [
400.00-1800.00]"/>
<cvParam cvRef="MS" accession="MS:1000616" name="preset scan configuration" value="3"/>
<cvParam cvRef="MS" accession="MS:1000803" name="analyzer scan offset" value="80"
unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
```

4.6 Element <binaryDataArray>

Definition: Data point arrays for default data arrays (m/z, intensity, time) and meta data arrays. Default data arrays MUST not have the attributes 'arrayLength' and 'dataProcessingRef'.

Type: dx:BinaryDataArrayType

Attributes:

Attribute Name	Data Type	Use	Definition
arrayLength	xs:nonNegativeInteger	optional	This optional attribute may override the 'defaultArrayLength' defined in SpectrumType. The two default arrays (m/z and intensity) should NEVER use this override option, and should therefore adhere to the 'defaultArrayLength' defined in SpectrumType. Parsing software can thus safely choose to ignore arrays of lengths different from the one defined in the 'defaultArrayLength' SpectrumType element.
dataProcessingRef	xs:IDREF	optional	This optional attribute may reference the 'id' attribute of the appropriate dataProcessing.
encodedLength	xs:nonNegativeInteger	required	The encoded length of the binary data array. Is always ZERO!

Subelements:

Subelement Name	min	max	Definition
referenceableParamGroupRef	0	unlim	A reference to a previously defined ParamGroup, which is a reusable container of one or more cvParams.
cvParam	0	unlim	This element holds additional data or annotation. Only controlled values are

			allowed here.
userParam	0	unlim	Uncontrolled user parameters (essentially allowing free text). Before using these, one should verify whether there is an appropriate CV term available, and if so, use the CV term instead
binary	1	1	Has to be always empty!

Example Context:

```
<binaryDataArray encodedLength="160" dataProcessingRef="XcaliburProcessing">
  <cvParam cvRef="MS" accession="MS:1000523" name="64-bit float" value=""/>
  <cvParam cvRef="MS" accession="MS:1000576" name="no compression" value=""/>
  <cvParam cvRef="MS" accession="MS:1000515" name="intensity array" value="" unitCvRef="MS"
unitAccession="MS:1000131" unitName="number of counts"/>
  <binary>AAAAAAAALkAAAAAAAAAsQAAAAAAAACpAAAAAAAAAKEAAAAAAA...</binary>
</binaryDataArray>
```

cvParam Mapping Rules:

Path mzML/run/spectrumList/spectrum/binaryDataArrayList/binaryDataArray

```
MUST supply a *child* term of MS:1000572 (binary data compression type) only once
  e.g.: MS:1000574 (zlib compression)
  e.g.: MS:1000576 (no compression)
MUST supply a *child* term of MS:1000513 (binary data array) only once
  e.g.: MS:1000514 (m/z array)
  e.g.: MS:1000515 (intensity array)
  e.g.: MS:1000516 (charge array)
  e.g.: MS:1000517 (signal to noise array)
  e.g.: MS:1000595 (time array)
  e.g.: MS:1000617 (wavelength array)
  e.g.: MS:1000786 (non-standard data array)
  e.g.: MS:1000820 (flow rate array)
  e.g.: MS:1000821 (pressure array)
  e.g.: MS:1000822 (temperature array)
MUST supply a *child* term of MS:1000518 (binary data type) only once
  e.g.: MS:1000521 (32-bit float)
  e.g.: MS:1000523 (64-bit float)
MUST supply term of IMS:1000103 (external array length) only once
MUST supply term of IMS:1000101 (external data) only once
MUST supply term of IMS:1000104 (external encoded length) only once
MUST supply term of IMS:1000102 (external offset) only once
```

Example cvParams:

```
<cvParam cvRef="MS" accession="MS:1000523" name="64-bit float" value=""/>
<cvParam cvRef="MS" accession="MS:1000576" name="no compression" value=""/>
<cvParam cvRef="MS" accession="MS:1000514" name="m/z array" value="" unitCvRef="MS"
unitAccession="MS:1000040" unitName="m/z"/>
<cvParam cvRef="MS" accession="MS:1000515" name="intensity array" value="" unitCvRef="MS"
unitAccession="MS:1000131" unitName="number of counts"/>
<cvParam cvRef="MS" accession="MS:1000595" name="time array" value="" unitCvRef="UO"
unitAccession="UO:0000010" unitName="second"/>
<cvParam cvRef="MS" accession="MS:1000521" name="32-bit float"/>
<cvParam cvRef="MS" accession="MS:1000574" name="zlib compression" value=""/>
```

Notes and Constraints: The arrayLength attribute need only be specified if it is different from the defaultArrayLength specified in the <spectrum> element. The arrayLength in imzML is zero by definition, therefore the arrayLength has to be defined in <spectrum>.

4.7 Element <binary>

Definition: The tag is always empty.

Type: none

Attributes: none

Subelements: none

Example Context: `<binary/>`

5. Conclusions

This document contains the specifications for using the imzML format to represent imaging mass spectrometry results, metadata and associated context. This specification, in conjunction with the XML Schema and the Reference Manual constitute a proposal for a standard from the partners of the Computis project.

6. Authors and contributors

Thorsten Schramm
Institute of Inorganic and Analytical Chemistry
Schubertstrasse 60, Haus 16,
35392 Giessen
Germany
thorsten.schramm@anorg.chemie.uni-giessen.de

Alfons Hester
Institute of Inorganic and Analytical Chemistry
Schubertstrasse 60, Haus 16,
35392 Giessen
Germany
alfons.hester@anorg.chemie.uni-giessen.de

Andreas Römpp
Institute of Inorganic and Analytical Chemistry
Schubertstrasse 60, Haus 16,
35392 Giessen
Germany
andreas.roempp@uni-giessen.de

The following people contributed to the model development, controlled vocabulary development, gave feedback or tested imzML:

- Ivo Klinkert, FOM Institute for Atomic and Molecular Physics (AMOLF), Amsterdam
- Ron M. Heeren, FOM Institute for Atomic and Molecular Physics (AMOLF), Amsterdam
- Markus Stoeckli, Novartis Institutes for BioMedical Research, Basel
- Jean Pierre Both, Commissariat à l'Énergie Atomique (CEA), Salclay
- Alain Brunelle, Centre National de la Recherche Scientifique (CNRS), Gif-sur-Yvette
- Bernhard Spengler, Justus Liebig University, Giessen

7. References

See the mzML version 1.1.0 documentation for further details.