



HAL
open science

imzML - A common data format for the flexible exchange and processing of mass spectrometry imaging data.

Thorsten Schramm, Alfons Hester, Ivo Klinkert, Jean-Pierre Both, Ron M A Heeren, Alain Brunelle, Olivier Lapr evote, Nicolas Desbenoit, Marie-France Robbe, Markus Stoeckli, et al.

► **To cite this version:**

Thorsten Schramm, Alfons Hester, Ivo Klinkert, Jean-Pierre Both, Ron M A Heeren, et al.. imzML - A common data format for the flexible exchange and processing of mass spectrometry imaging data.. Journal of Proteomics, 2012, 75 (16), pp.5106-10. 10.1016/j.jprot.2012.07.026 . hal-00741330

HAL Id: hal-00741330

<https://hal.science/hal-00741330v1>

Submitted on 17 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

Manuscript Number: JPROT-D-12-00290R1

Title: imzML - a common data format for the flexible exchange and processing of mass spectrometry imaging data

Article Type: SI: Mass Spectrometry Imaging

Section/Category: Technical Note

Keywords: mass spectrometry imaging
data format
data processing
standardization

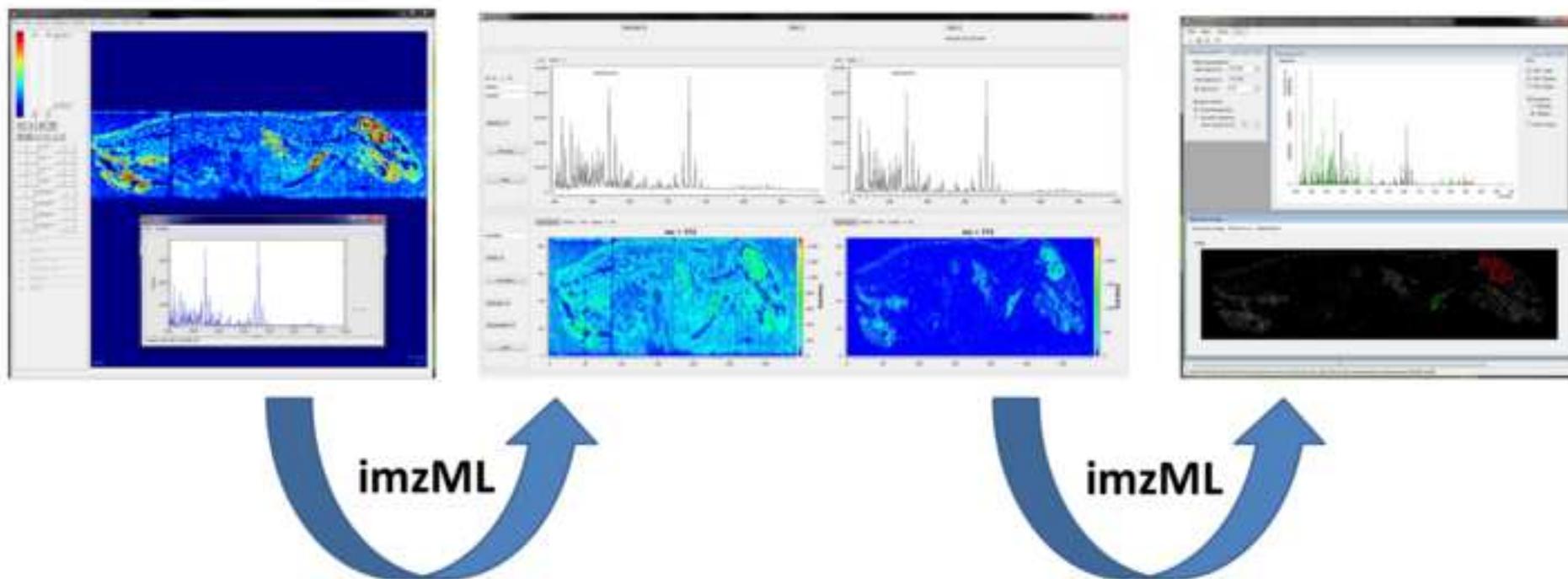
Corresponding Author: Dr Andreas Roempp, PhD

Corresponding Author's Institution: Justus Liebig University

First Author: Thorsten Schramm

Order of Authors: Thorsten Schramm; Alfons Hester; Ivo Klinkert; Jean-Pierre Both; Ron M Heeren; Alain Brunelle; Olivier Laprevote; Marie-France Robbe; Markus Stoeckli; Bernhard Spengler; Andreas Roempp, PhD

Abstract: The application of MS imaging is rapidly growing with a constantly increasing number of different instrumental systems and software tools. The data format imzML was developed to allow the flexible and efficient exchange of MS imaging data between different instruments and data analysis software. imzML data is divided in two files which are linked by a universally unique identifier (UUID). Experimental details are stored in an XML file which is based on the HUPO-PSI format mzML. Information is provided in the form of a 'controlled vocabulary' (CV) in order to unequivocally describe the parameters and to avoid redundancy in nomenclature. Mass spectral data are stored in a binary file in order to allow efficient storage. imzML is supported by a growing number of software tools. Users are no longer limited to proprietary software, but are able to use the processing software best suited for a specific question or application. MS imaging data from different instruments can be converted to imzML and displayed with identical parameters in one software package for easier comparison. All technical details necessary to implement imzML and additional background information is available at www.imzml.org.



Highlights

- Common format for exchange of mass spectrometry imaging data
- Comprehensive and efficient description of MS imaging experiments
- Converters and data analysis tools are available
- Flexible structure allows adaption to new methods and instruments

1 imzML – a common data format for the flexible exchange and processing of mass
2 spectrometry imaging data

3 Thorsten Schramm¹, Alfons Hester¹, Ivo Klinkert², Jean-Pierre Both³, Ron M. A. Heeren²,
4 Alain Brunelle⁴, Olivier Laprévote⁵, Nicolas Desbenoit¹, Marie-France Robbe³, Markus
5 Stoeckli⁶, Bernhard Spengler¹ and Andreas Römpf^{1*}

6 ¹Justus Liebig University, Giessen, Germany; ²FOM Institute AMOLF, Amsterdam,
7 Netherlands; ³CEA-LIST, Saclay, France; ⁴Centre de recherche de Gif, Institut de Chimie des
8 Substances Naturelles, CNRS, Gif-sur-Yvette Cedex, France; ⁵Chimie et Toxicologie
9 Analytique et Cellulaire, Université Paris Descartes, Paris, France; ⁶Novartis Institutes for
10 BioMedical Research, Basel, Switzerland

11

12 * Corresponding author

13 Dr. Andreas Roempp
14 Institute of Inorganic and Analytical Chemistry
15 - Analytical Chemistry -
16 Justus Liebig University Giessen
17 Schubertstrasse 60, Build. 16
18 D-35392 Giessen, Germany
19 phone: +49-641-99 34802
20 fax: +49-641-99 34809
21 email: Andreas.Roempp@anorg.Chemie.uni-giessen.de

22

23 **Abstract**

24 The application of mass spectrometry imaging (MS imaging) is rapidly growing with a
25 constantly increasing number of different instrumental systems and software tools. The data
26 format imzML was developed to allow the flexible and efficient exchange of MS imaging
27 data between different instruments and data analysis software. imzML data is divided in two
28 files which are linked by a universally unique identifier (UUID). Experimental details are
29 stored in an XML file which is based on the HUPO-PSI format mzML. Information is
30 provided in the form of a ‘controlled vocabulary’ (CV) in order to unequivocally describe the
31 parameters and to avoid redundancy in nomenclature. Mass spectral data are stored in a binary
32 file in order to allow efficient storage. imzML is supported by a growing number of software
33 tools. Users will be no longer limited to proprietary software, but are able to use the
34 processing software best suited for a specific question or application. MS imaging data from
35 different instruments can be converted to imzML and displayed with identical parameters in
36 one software package for easier comparison. All technical details necessary to implement
37 imzML and additional background information is available at www.imzml.org.

38

40 **Introduction**

41 In mass spectrometry imaging (MS Imaging) a sample of interest is scanned and the resulting
42 ion signals are reconstructed into a pixel image. Several overviews of this method have been
43 published recently [1, 2]. The application of MS imaging is rapidly growing with a constantly
44 increasing number of different instrumental systems and software tools. This results in a need
45 of exchangeability of MS imaging data between different instruments and data analysis
46 software. As for many analytical methods, data processing has become an essential part of the
47 workflow for MS imaging. MS Imaging data consist of several thousand spectra which
48 frequently result in data files of several gigabytes. Mass spectra of one experiment are usually
49 acquired with identical measurement parameters. Existing standards such as the DICOM
50 standard [3] for in-vivo imaging data or the mzML standard [4] established by HUPO-PSI [5]
51 are not suitable to adequately represent an MS imaging experiment. Therefore a common data
52 format dedicated to mass spectrometry imaging data was developed within the EU project
53 COMPUTIS (www.computis.org). The data format is in part based on the HUPO-PSI format
54 mzML (see details below) and is thus called imzML for ‘imaging mzML’.

55 This technical report provides an overview of the most important properties and features of
56 the imzML format. Additional information and all technical details necessary to implement
57 imzML is provided on the website **www.imzml.org** and in a recently published book chapter
58 [6].

59 **Methodology**

60 MS imaging data in imzML is divided into two separate files (Figure 1). This structure,
61 consisting of a small file (text or XML) and a larger (binary) file, has been shown to be
62 efficient in previous data formats for MS imaging, for example in BioMap [7] and internal
63 data formats at FOM Institute and Justus Liebig University. All metadata (e.g. instrumental
64 parameters, sample details) are stored in an XML file. Mass spectral data is stored in a binary
65 file to ensure efficient storage. Corresponding XML and binary files contain a universally
66 unique identifier (UUID) [8] in order to link them unequivocally and to prevent loss of
67 information.

68 The **XML file** (*.imzml) contains all relevant information about the MS imaging experiment.
69 In order to stay as close as possible to existing formats this metadata file is based on the mass

70 spectrometry data standard mzML [4] developed by HUPO-PSI [9]. Discussions with HUPO-
71 PSI resulted in the strategy to maintain imzML as a separate format, but keep the XML
72 structure consistent with mzML. Information is provided in the form of a ‘controlled
73 vocabulary’ (CV) which is stored in an open biomedical ontology [10]. This approach is used
74 in order to unequivocally describe the parameters and to avoid redundancy in nomenclature.
75 The mzML controlled vocabulary [11] has been supplemented by parameters which are
76 necessary for a comprehensive description of MS imaging experiments. These additional
77 imzML CV parameters (including x/y position, scan direction/pattern, pixel size) are stored in
78 the imagingMS.obo file [12].

79 The **imaging binary data file** (*.ibd) contains the mass spectral data of the MS imaging
80 measurement. In order to ensure efficient storage, two different binary modes are defined:
81 continuous and processed. ‘Continuous’ means that an intensity value is stored for each m/z
82 bin even if there is no measurable signal (resulting in an intensity of zero). This data structure
83 is used for many MALDI-TOF mass spectrometers. As a result the m/z axis is identical for all
84 mass spectra of one image. Therefore it is sufficient to store the m/z array only once in the
85 binary file. This structure can reduce the file size significantly (up to a factor of 2). On the
86 other hand mass spectra are often processed before they are stored e.g. for noise-reduction,
87 peak-picking, deisotoping. This results in discontinuous and non-constant m/z arrays. In this
88 case the m/z array has to be stored for each spectrum separately. This data structure was
89 termed ‘Processed’.

90 The two files (XML and binary) are connected by offset values in the XML file that indicate
91 the position of the corresponding data in the binary file. This allows for fast reading/access of
92 large data sets. The combination of non-redundant metadata representation and binary data
93 leads to efficient storage. The resulting datasets are comparable in size to the proprietary data
94 and the separate metadata file allows flexible handling of large datasets. This is important
95 since MS imaging experiments typically consist of several thousand spectra, and increased
96 file size due to format conversion could lead to data sets which are very difficult to handle.

97 imzML has been extensively discussed with academic and industrial users at various
98 occasions in recent years. The current version 1.1.0 of imzML was released at the
99 International Mass Spectrometry Conference (IMSC) in Bremen, Germany in September
100 2009. The flexible structure based on XML and the controlled vocabulary allows for
101 compatibility with new instruments and methods in the future. Technical documentation

102 including XML scheme, mapping file, controlled vocabulary (obo file) and example files are
103 provided on www.imzml.org.

104

105 **Implementation and Discussion**

106 A number of software tools already support imzML. This allows for choosing from a
107 (growing) number of options to display and process MS imaging data. Users will be no longer
108 limited to proprietary software, but are able to use the processing software best suited for a
109 specific question or application. Multiple software tools can also be combined in one
110 workflow using imzML as demonstrated in Figure 2. In this example a whole body rat section
111 was coated with α -Cyano-4-hydroxycinnamic acid (CHCA) and analysed with a QStar mass
112 spectrometer (AB Sciex). Mass spectra were acquired for 344 x 87 pixels (500 μm pixel size)
113 in the mass range $m/z = 300\text{-}1000$ u. Data was saved in the Analyze 7.5 data format and
114 opened in 'BioMap' for initial inspection (Figure 2A). Subsequently the dataset was
115 converted to imzML using the 'Toimzml' converter developed by the CEA-LIST institute.
116 The imzML file was opened in 'SpectViewer' (developed by CEA-LIST) and a wavelet
117 denoising procedure was applied in order to improve image quality (Figure 2B). This
118 particular processing function is not available in the other software tools. The processed data
119 was saved in a new imzML file which was subjected to further analysis.

120 The 'Datacube Explorer' developed by AMOLF allows for fast browsing of MS images and
121 can thus be used to visually search for signals that show interesting spatial features. The
122 denoised imzML data set was displayed in Datacube Explorer and spectra from two regions of
123 interest (ROI) were extracted (Figure 2C). The data analysis tool 'Mirion' developed by JLU
124 can be used to semiautomatically generate MS images based on parameters such as pixel
125 coverage or incremental mass. It can also be used to open multiple files as shown in Figure
126 2D (original and denoised imzML data set), data from these files can be combined for
127 example by RGB overlay (not shown). The combination of processing steps within the four
128 programmes is only possible due to the standardised and open data format imzML. A detailed
129 description of the different software packages would be beyond the scope of this technical
130 note. The purpose of the example shown in Figure 2 is to demonstrate the possibility to open
131 and process imzML files in different MS imaging software packages. Details about
132 functionalities of SpectViewer, Datacube Explorer and Mirion will be provided in separate
133 publications which are currently in preparation.

134 The demonstrated workflow is only one example for the use of imzML for analysis of MSI
135 data with different tools. Another application is the comparison of measurements from
136 different instruments. Options for data processing (e.g. binning, normalization) and displaying
137 (e.g. interpolation options, color schemes) vary strongly between different MS imaging
138 software tools and can make a direct comparison of data difficult. This problem can be
139 avoided, if all data sets are converted to imzML and the data is displayed with identical
140 parameters in one software package. Examples of imzML data sets from different instrument
141 platforms are shown in Figure 3. All images are displayed in the same software package (two
142 masses in green and red, no interpolation). Data from the most commonly used matrix-
143 assisted desorption/ionization (MALDI) mass spectrometers (Thermo, Bruker, AB Sciex and
144 Waters) are shown in Figures 3A-D. A data set originating from secondary ion mass
145 spectrometry (SIMS) is shown in Figure 3E. The majority of MS imaging applications are
146 based on MALDI and SIMS, but other ionization techniques are increasingly used. Data
147 acquisition in desorption electrospray ionization (DESI) mode is based on one data file per
148 line scan and conversion to imzML therefore differs from most MALDI experiments where
149 usually a single file is acquired per image. The ‘RAW to imzML’ converter developed by
150 Justus Liebig University was modified to convert multiple line scans into a unified imzML
151 file. An example of a DESI imzML file generated with this converter is shown in Figure 3F.
152 This conversion tool has also been used in previous studies in combination with the Datacube
153 Explorer for analyzing and displaying MS imaging data[13, 14]. More information on sample
154 details, image properties and conversion method are provided in the ‘imzML gallery’ on
155 www.imzml.org. This page also includes additional examples of MS images generated with
156 imzML.

157 The examples discussed above illustrate the functionalities and possibilities of the imzML
158 data format. A number of additional software tools are currently adopted for usage with
159 imzML. This includes ‘MSImageView’ the successor of BioMap which is probably the most
160 widely used independent software to analyze MS imaging data. The first two commercial
161 software packages supporting imzML have been released recently [15, 16]. The concept of
162 imzML is also actively supported by major vendors of instrumentation for MS imaging
163 including Waters Corporation (Manchester, United Kingdom), Thermo Fisher Scientific
164 (Bremen, Germany) and Bruker Daltonik (Bremen, Germany). Their data formats are
165 integrated in the imzML workflow through export options in proprietary software suites
166 (Waters) or by external converters being developed by the MS imaging community with
167 technical support from the companies (Thermo, Bruker). An updated list of available tools

168 and workflows can be found on the imzML website. imzML was chosen as the central
169 platform for exchange of data in the European COST network ‘Mass Spectrometry Imaging:
170 New Tools for Healthcare Research’ which includes more than 30 MS imaging groups [17].
171 The future development of imzML-based software will also be coordinated in this framework.
172 These activities ensure a wide distribution and active development of imzML as the data
173 standard for mass spectrometry imaging data.

174 **Conclusions**

175 The data format imzML enables efficient and flexible exchange of MS imaging data. It allows
176 for a flexible selection of data analysis tools. imzML is increasingly used in the MS imaging
177 community. A number of software tools are available and many more are currently being
178 adapted to imzML. A common data format is not a mere technical detail, but has significant
179 impact on practical work: fully searchable mass spectrometry imaging data can be shared with
180 collaborators in biological or clinical labs without being restricted to proprietary vendor
181 software. imzML is thus an important step towards more efficient collaboration as well as
182 more flexible and transparent data processing in mass spectrometry imaging. These are key
183 requirements for MS imaging to finally be established as a routine method.

184

185 **Acknowledgements**

186 First and foremost we would like to thank all the members of the HUPO Proteomics
187 Standards Initiative (PSI) for developing the data format mzML which has been used as a
188 model for imzML. We are especially thankful to Lennart Martens, Eric Deutsch, Randall
189 Julian and Martin Eisenacher for many helpful discussions in the last years. We would like to
190 thank Sabine Guenther, Corinna Henkel, Brendan Prideaux, Emmanuelle Claude, Zoltan
191 Takats for providing MS imaging data sets. This work was supported by the European Union
192 (Contract LSHG-CT-2005-518194 COMPUTIS). JLU acknowledges support by the Hessian
193 Ministry of Science and Art (LOEWE focus Ambiprobe).

194

195 **References**

- 196
- 197 [1] Chughtai K, Heeren RMA. Mass Spectrometric Imaging for Biomedical Tissue Analysis. Chem Rev. 2010;110:3237-77.
- 198
- 199 [2] McDonnell LA, Heeren RMA. Imaging mass spectrometry. Mass Spectrometry Reviews. 2007;26:606-43.
- 200
- 201 [3] DICOM. Digital imaging and communications in medicine - DICOM. <http://medical.nema.org>.
- 202 [4] Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Rompp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz P-A, Deutsch EW. mzML - a Community Standard for Mass Spectrometry Data. Molecular & Cellular Proteomics. 2011;10.
- 203
- 204 [5] Hermjakob H. The HUPO proteomics standards initiative - Overcoming the fragmentation of proteomics data. Proteomics. 2006;34-8.
- 205
- 206 [6] Römpf A, Schramm T, Hester A, Klinkert I, Heeren R, Stöckli M, Both J-P, Brunelle A, Spengler B. Imaging mzML (imzML) – a common data format for imaging mass spectrometry. In: Hamacher M, Stephan C, Eisenacher M, editors. Data Mining in Proteomics. New York: Humana Press; 2010. p. 205-24.
- 207
- 208 [7] Biomap. http://maldi-msi.org/index.php?option=com_content&task=view&id=14&Itemid=39.
- 209 [8] Network_Working_Group. RFC 4122 - A Universally Unique Identifier (UUID) URN Namespace. <http://tools.ietf.org/html/rfc4122>.
- 210
- 211 [9] HUPO-PSI. The Human Proteome Organization: Proteomics Standards Initiative. <http://psidev.info>.
- 212 [10] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S, Consortium OBI. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. 2007;25:1251-5.
- 213 [11] HUPO-PSI. Mass Spectrometry Standards Working Group - controlled vocabulary. <http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo>.
- 214 [12] MS Imaging controlled vocabulary. http://www.maldi-msi.org/index.php?option=com_content&view=article&id=193&Itemid=66.
- 215 [13] Li B, Bjarnholt N, Hansen SH, Janfelt C. Characterization of barley leaf tissue using direct and indirect desorption electrospray ionization imaging mass spectrometry. J Mass Spectrom. 2011;46:1241-6.
- 216 [14] Thunig J, Hansen SH, Janfelt C. Analysis of Secondary Plant Metabolites by Indirect Desorption Electrospray Ionization Imaging Mass Spectrometry. Anal Chem. 2011;83:3256-9.
- 217 [15] PREMIERBiosoft. MALDIVision. <http://www.premierbiosoft.com/maldi-tissue-imaging/index.html>.
- 218 [16] Imabiotech. Quantinetix MALDI Imaging Software. <http://www.imabiotech.com/Quantinetix-TM-Maldi-Imaging.html?lang=en>.
- 219 [17] McDonnell LA, Heeren RMA, Andrén PE, Stoeckli M, Corthals GL. Going forward: Increasing the accessibility of imaging mass spectrometry. Journal of Proteomics.
- 220
- 221
- 222
- 223
- 224
- 225
- 226
- 227
- 228
- 229
- 230
- 231
- 232
- 233
- 234
- 235

236

237

238 Figure 1: Scheme of imzML structure (adapted from www.imzml.org).

239

240

241 Figure 2: Example of sequential processing of an MS imaging data set based on imzML. MS imaging
242 data acquired on an AB Sciex QStar instrument was converted to imzML. The imzML file was opened
243 and processed in different software packages for analysis of MS imaging data. A) BioMap: Selected
244 ion image of $m/z = 772$ and averaged mass spectrum, B) SpectViewer: averaged mass spectrum and
245 selected ion image of $m/z = 772$ for original data (left) and after application of wavelet denoising
246 algorithm (right), C) Datacube Explorer: mass spectrum of full image and two regions of interest (ROI)
247 which are indicated in selected ion image of $m/z 772$, D) Mirion: selected ion image of $m/z = 772$
248 from original (top) and denoised (bottom) data set.

249

250 Figure 3: Examples of data sets from different instrument platforms that were converted to imzML.
251 All images are displayed in the same software tool (Mirion) with identical settings (showing an
252 overlay of two masses in red and green, no interpolation). Ionization type and mass spectrometer
253 type are indicated for each measurement. A) mouse kidney, MALDI, LTQ Orbitrap (Thermo), data
254 provided by Sabine Guenther (Justus Liebig University, Giessen, Germany), B) mouse brain (left half),
255 MALDI, ultrafleXtreme (Bruker), data provided by Corinna Henkel (Bochum University, Germany), C)
256 whole body rat section (head), MALDI, QStar (ABSciex), same data as shown in Figure 2, D) whole bod
257 yrat section, MALDI, Synapt G2 (Waters), data provided by Emmanuelle Claude (Waters, Manchester,
258 UK), E) rabbit eye, SIMS, TOF-SIMS IV (IonTOF), data provided by Alain Brunelle (CNRS, Gif-sur-
259 Yvette, France), F) tumor tissue, DESI, Exactive (Thermo), data provided by Zoltan Takats (Imperial
260 College, London, UK). More details on these data sets can be found in the 'imzML gallery' on the
261 www.imzml.org.

262

Figure 1
[Click here to download high resolution image](#)

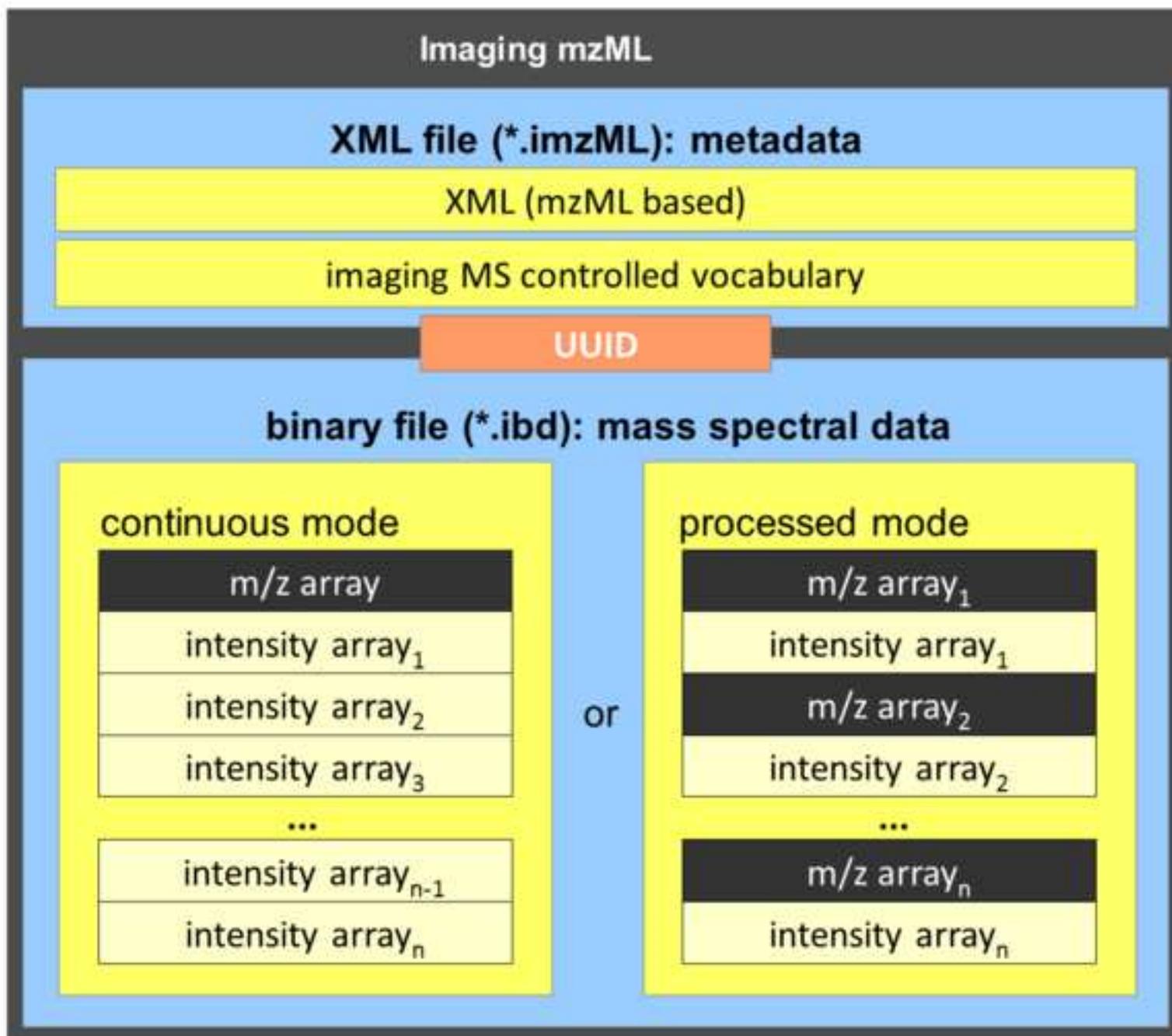


Figure 2
[Click here to download high resolution image](#)

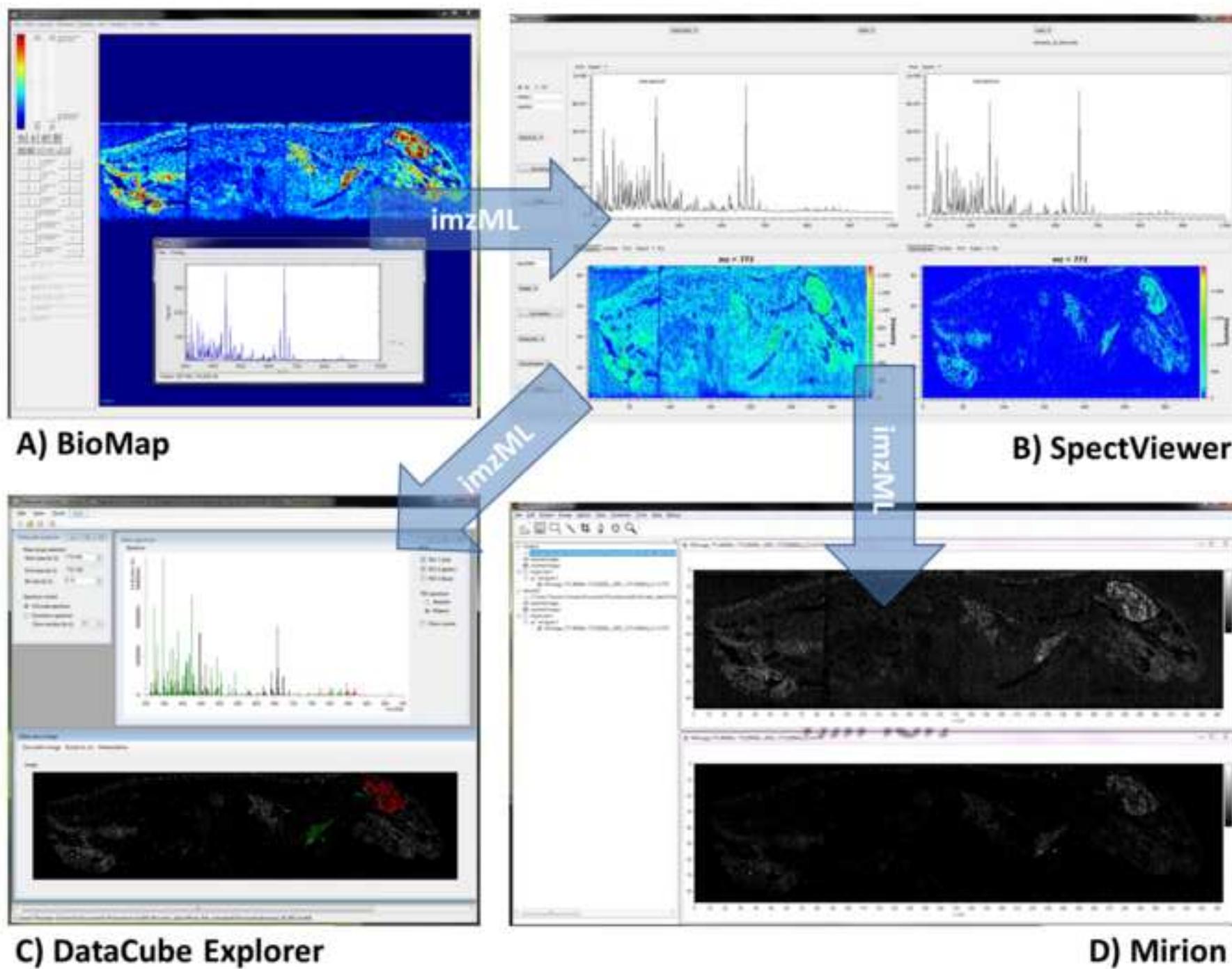


Figure 3
[Click here to download high resolution image](#)

