mzXML

Patrick McConnell
Duke Bioinformatics Shared Resource
Duke Comprehensive Cancer Center
patrick.mcconnell@duke.edu

Agenda

- Schema
- Example XML
- Data discussion
- Tools
- Databases
- Datasets

Schema Elements

- Instrumentation
- Data processing
- Separation technique
- Spot description
- m/z scan values
- MALDI acquisition
- Data integrity

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

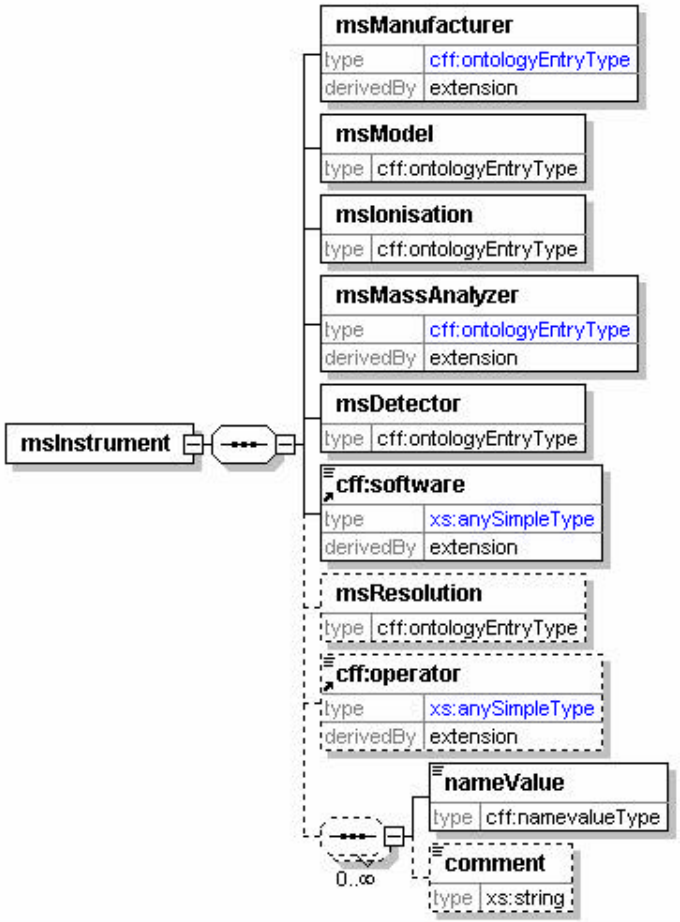
File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_msInstrument_Link03314120 Go Links

element **msRun/msInstrument**

diagram



namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1
children	**msManufacturer msModel msIonisation msMassAnalyzer msDetector software msResolution operator nameValue comment**
annotation	documentation General information about the MS instrument.
source	<xs:element name="msInstrument" minOccurs="0">

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

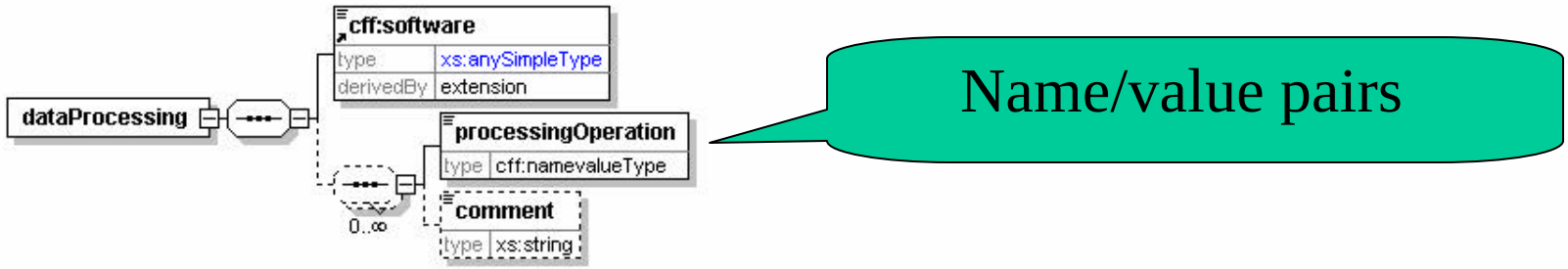
File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_dataProcessing_Link0346AEB8 Go Links

element **msRun/dataProcessing**

diagram



namespace http://sashimi.sourceforge.net/schema_revision/mzXML_2.1

children **software processingOperation comment**

attributes

Name	Type	Use	Default	Fixed	Annotation
intensityCutoff	xs:float	optional			documentation Intensity threshold for inclu
centroided	xs:boolean	optional			documentation Is the data centroided
deisotoped	xs:boolean	optional			documentation Is the data deisotoped
chargeDeconvoluted	xs:boolean	optional			documentation Has the charge state been
spotIntegration	xs:boolean	optional			

annotation documentation Description of any manipulation (from the first conversion to mzXML format until the creation of the current mzXML instance document) applied to the data.

source

```
<xs:element name="dataProcessing" maxOccurs="unbounded">
 <xs:annotation>
  <xs:documentation>Description of any manipulation (from the first conversion to mzXML format until the creation of the current mzXML instance document) applied to the
 </xs:annotation>
 <xs:complexType>
  <xs:sequence>
   <xs:element ref="cff:software">
    <xs:annotation>
     <xs:documentation>Software used to convert the data. If data has been processed (e.g. profile > centroid) by any additional progs these should be added too.</xs:doc
    </xs:annotation>
   </xs:element>
   <xs:sequence minOccurs="0" maxOccurs="unbounded">
    <xs:element name="processingOperation" type="cff:namevalueType">
     <xs:annotation>
      <xs:documentation>Any additional manipulation not included elsewhere in the dataProcessing element.</xs:documentation>
     </xs:annotation>
    </xs:element>
    <xs:element name="comment" type="xs:string" minOccurs="0">
```

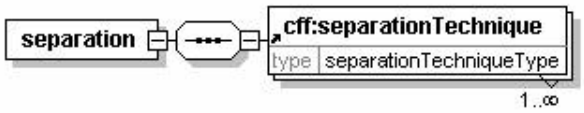
http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_separation_Link03355B48 Go Links

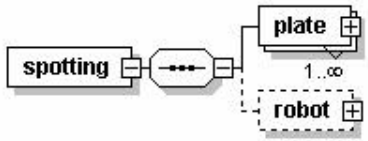
element **msRun/separation**

diagram	
namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1
children	**separationTechnique**
annotation	documentation Information about the separation technique, if any, used right before the acquisition.

source	`<xs:element name="separation" minOccurs="0">`
	`<xs:annotation>`
	`<xs:documentation>Information about the separation technique, if any, used right before the acquisition.</xs:documentation>`
	`</xs:annotation>`
	`<xs:complexType>`
	`<xs:sequence>`
	`<xs:element ref="cff:separationTechnique" maxOccurs="unbounded"/>`
	`</xs:sequence>`
	`</xs:complexType>`
	`</xs:element>`

Externally defined

element **msRun/spotting**

diagram	
namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1
children	**plate robot**
annotation	documentation Acquisition independent properties of a MALDI experiment.

source	`<xs:element name="spotting" minOccurs="0">`
	`<xs:annotation>`
	`<xs:documentation>Acquisition independent properties of a MALDI experiment.</xs:documentation>`
	`</xs:annotation>`
	`<xs:complexType>`

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

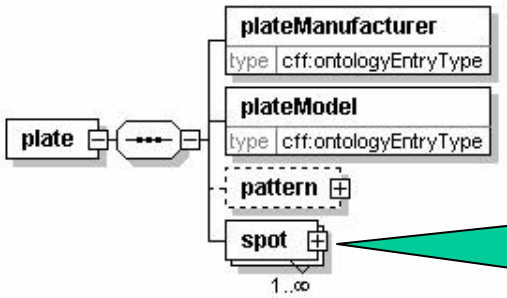
File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_plate_Link033548B0 Go Links

element **msRun/spotting/plate**

diagram



spotID
spotXPosition
spotYPosition
spotDiameter

namespace: http://sashimi.sourceforge.net/schema_revision/mzXML_2.1

children: **plateManufacturer** **plateModel** **pattern** **spot**

attributes

Name	Type	Use	Default	Fixed	Annotation
plateID	xs:positiveInteger	required			documentation Unique identifier for the plate.
spotXCount	xs:positiveInteger	required			
spotYCount	xs:positiveInteger	required			

identity constraints

	Name	Refer	Selector	Field(s)
key	plateKey		.	@plateID

annotation: documentation Information about a MALDI plate.

source

```
<xs:element name="plate" maxOccurs="unbounded">
 <xs:annotation>
  <xs:documentation>Information about a MALDI plate.</xs:documentation>
 </xs:annotation>
 <xs:complexType>
  <xs:sequence>
   <xs:element name="plateManufacturer" type="cff:ontologyEntryType"/>
   <xs:element name="plateModel" type="cff:ontologyEntryType"/>
   <xs:element name="pattern" minOccurs="0">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="spottingPattern" type="cff:ontologyEntryType"/>
      <xs:element name="orientation">
       <xs:annotation>
        <xs:documentation>Defines the orientation of the spotting pattern, with the IDs of the fist two consecutive spotted spots.</xs:documentation>
       </xs:annotation>
```

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

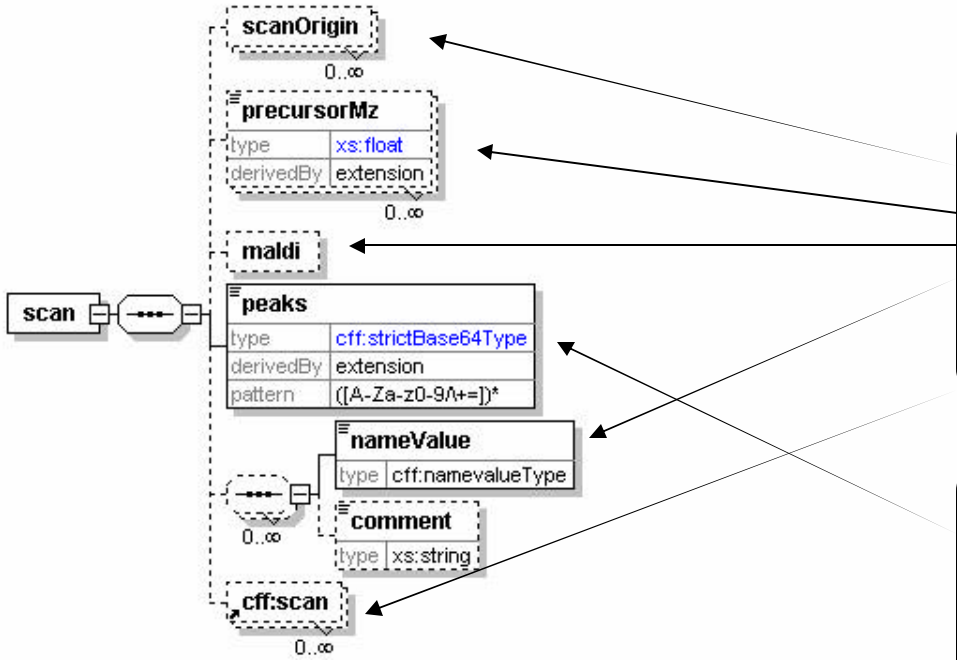
File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_scan_Link03348B30 Go Links

element **scan**

diagram



Metadata

Let's take a closer look at real data

namespace http://sashimi.sourceforge.net/schema_revision/mzXML_2.1

children **scanOrigin precursorMz maldi peaks nameValue comment scan**

used by elements **msRun scan**

attributes

Name	Type	Use	Default	Fixed	Annotation	
num	xs:positiveInteger	required			documentation	Scan number
msLevel	xs:positiveInteger	required			documentation	1: MS 2: MS/MS ...
peaksCount	xs:nonNegativeInteger	required			documentation	Total number of m/z-intensity pairs in the scan
polarity	xs:string	optional			documentation	+: positive -: negative any
scanType	xs:string	optional				
centroided	xs:boolean	optional			documentation	This only maters if centroided attribute of dataProcessing element is 0 or absent.

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_peaks_Link033546C0 Go Links

element **scan/peaks**

diagram

peaks

type	cff:strictBase64Type
derivedBy	extension
pattern	([A-Za-z0-9/\+=])*

Binary!

namespace: http://sashimi.sourceforge.net/schema_revision/mzXML_2.1

type: extension of **strictBase64Type**

facets: pattern ([A-Za-z0-9/\+=])*

attributes

Name	Type	Use	Default	Fixed	Annotation
precision	xs:positiveInteger	required			documentation nr of bits used by each component
byteOrder	xs:string	required		network	documentation Byte order of the encoded binary information (must be network).
pairOrder	xs:string	required		m/z-int	documentation Order of the m/z intensity pairs (must be m/z-int).

annotation: documentation This is the actual data encoded using base64. Byte order must be network. The order of the pairs must be m/z – intensity.

source

```
<xs:element name="peaks" nillable="1">
 <xs:annotation>
  <xs:documentation>This is the actual data encoded using base64. Byte order must be network. The order of the pairs must be m/z – intensity.</xs:documentation>
 </xs:annotation>
 <xs:complexType>
  <xs:simpleContent>
   <xs:extension base="cff:strictBase64Type">
    <xs:attribute name="precision" use="required">
     <xs:annotation>
      <xs:documentation>nr of bits used by each component</xs:documentation>
     </xs:annotation>
     <xs:simpleType>
      <xs:restriction base="xs:positiveInteger">
       <xs:enumeration value="32"/>
       <xs:enumeration value="64"/>
      </xs:restriction>
     </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="byteOrder" type="xs:string" use="required" fixed="network">
     <xs:annotation>
      <xs:documentation>Byte order of the encoded binary information (must be network).</xs:documentation>
     </xs:annotation>
    </xs:attribute>
```

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_maldi_Link0334D838 Go Links

element **scan/maldi**

More nifty meta-data

diagram	**maldi**
namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1

attributes

Name	Type	Use	Default	Fixed	Annotation
plateID	xs:string	required			documentation Unique plate ID. This ID must have been defined also in the plate element.
spotID	xs:string	required			documentation Unique spot ID. This ID must have been defined also in the spot element.
laserShootCount	xs:positiveInteger	optional			documentation Number of times the laser was fired to generate this scan.
laserFrequency	xs:duration	optional			documentation Frequency of the laser.
laserIntensity	xs:positiveInteger	optional			documentation Intensity of laser.
collisionGas	xs:boolean	optional			

identity constraints

	Name	Refer	Selector	Field(s)
keyref	onlyValidPlateID	cff:plateKey	.	@plateID
keyref	onlyValidSpotID	cff:spotKey	.	@spotID

annotation: documentation Acquisition dependent properties of a MALDI experiment.

source

```
<xs:element name="maldi" minOccurs="0">
 <xs:annotation>
  <xs:documentation>Acquisition dependent properties of a MALDI experiment.</xs:documentation>
 </xs:annotation>
 <xs:complexType>
  <xs:attribute name="plateID" type="xs:string" use="required">
   <xs:annotation>
    <xs:documentation>Unique plate ID. This ID must have been defined also in the plate element.</xs:documentation>
   </xs:annotation>
  </xs:attribute>
  <xs:attribute name="spotID" type="xs:string" use="required">
   <xs:annotation>
    <xs:documentation>Unique spot ID. This ID must have been defined also in the spot element.</xs:documentation>
   </xs:annotation>
  </xs:attribute>
  <xs:attribute name="laserShootCount" type="xs:positiveInteger" use="optional">
   <xs:annotation>
    <xs:documentation>Number of times the laser was fired to generate this scan.</xs:documentation>
   </xs:annotation>
  </xs:attribute>
  <xs:attribute name="laserFrequency" type="xs:duration" use="optional">
   <xs:annotation>
    <xs:documentation>Frequency of the laser.</xs:documentation>
```

http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1.html#element_sha1_Link0334ED38 Go Links

element **msRun/sha1**

diagram	**sha1** type: xs:string derivedBy: restriction length: 40
namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1
type	restriction of **xs:string**
facets	length 40
annotation	documentation sha-1 sum for this file (from the beginning of the file up to (and including) the opening tag of sha1

source

```
<xs:element name="sha1" minOccurs="0">
 <xs:annotation>
  <xs:documentation>sha-1 sum for this file (from the beginning of the file up to (and including) the opening tag of sha1</xs:documentation>
 </xs:annotation>
 <xs:simpleType>
  <xs:restriction base="xs:string">
   <xs:length value="40"/>
  </xs:restriction>
 </xs:simpleType>
</xs:element>
```

Data integrity!

element **operator**

diagram	**operator** type: xs:anySimpleType derivedBy: extension
namespace	http://sashimi.sourceforge.net/schema_revision/mzXML_2.1
type	extension of **xs:anySimpleType**
used by	element **msRun/msInstrument**

attributes

Name	Type	Use	Default	Fixed	Annotation
first	xs:string	required			documentation First name
last	xs:string	required			documentation Last name
phone	xs:string	optional			documentation Phone number

Example

- 7 protein mix:
  - Rabbit glycogen phosphorylase
  - E. Coli Beta-galactosidase
  - Bovine serum albumin
  - Myosin
  - Chicken Ovalbumin
  - Bovine serotransferrin

File:Lines
000.mzXML:73114
020.mzXML:70618
040.mzXML:68746
060.mzXML:69838
080.mzXML:66874
100.mzXML:67498
150.mzXML:66640
200.mzXML:66854
300.mzXML:66952
500.mzXML:66562
900.mzXML:66172

136,796,000 bytes of XML

2:mendel.duhs.duke.edu - [default] - F-Secure SSH Client

File Edit View Window Help

Quick Connect Profiles


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<msRun
 xmlns="http://sashimi.sourceforge.net/schema/"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://sashimi.sourceforge.net/schema/ http://sashimi.sourceforge.net/schema/MsXML.xsd"
 scanCount="3748"
 startTime="PT0.440000S"
 endTime="PT5102.550000S">
  <parentFile fileName="000.mzXML"
              fileType="RAWData"
              fileSha1="da39a3ee5e6b4b0d3255bfef95601890afd80709"/>
  <instrument manufacturer="ThermoFinnigan"
              model="LCQ Classic"
              ionisation="ESI"
              msType="Ion Trap">
    <software type="acquisition"
              name="Xcalibur"
              version="1.3"/>
  </instrument>
  <dataProcessing>
    <software type="conversion"
              name="ReAdW"
               version="0.7.3"/>
  </dataProcessing>
  <scan num="1"
        msLevel="1"
        peaksCount="992"
        polarity="+"
        scanType="full"
        centroided="1"
        retentionTime="PT0.440000S"
        startMz="300.0000"
        endMz="1500.0000"
        lowMz="300.5052"
        highMz="1499.0991"
        basePeakMz="391.0454"
        basePeakIntensity="6986078.0000"
        totIonCurrent="130127006.0000">
```


000.mzXML (0%)

Metadata

Data is found in the scan element

Connected to mendel.duhs.duke.edu | SSH2 - 3des-cbc - hmac-md5 - none | 124x39 | 15, 39 | 00:08:24

2:mendel.duhs.duke.edu - [default] - F-Secure SSH Client

File Edit View Window Help

Quick Connect Profiles


```
<peaks precision="32">Q5ZAqkgTwkBDlrcORq1gAEOXDEpIvL7AQ5dvzkgHKoBDl/D6SOfXIEOYhRxIh8jgQ5j9xEjOJEBDmZ4ASJE9IEOaGKRIDQLAQ5
p/4EkQfpBDmz3oSMHSwEOb6mJIM5SAQ5xzNEjnvmBDnQ+ESD2wwEOdi8pIUhtAQ54A1EjCTkBDnmYcSJXywEOfF7xIbZ/AQ593PEid7oBDn/COSCPmAEOgZcRIQV
OAQ6DgaEix/OBDoVkaSA9RQEOhrzZIjSRgQ6IatkhBfwBDopaAR2HdAEOi9B5IlY8AQ6NbMEfqbIBDo+wWSHY6wEOkk/hH4riAQ6UTykgLEUBDpYpSSGVcgEOmJ4
xIh/zAQ6artEhuX4BDpyb8SIFNQEOnowBIyT6AQ6gM8EgPDABDqGaySJ+34EOowjhGpTAAQ6ksQkgsIMBDqbRgR/6wgEOqeUxIZrgAQ6raNEfELQBDqOcCSToQOE
Or4qZI2CAAQ6yEFkggNQBDrPdYR7OpAEOtYhZHlN4AQ631qEeZaIBDrlpgSJOpwEOvBOxIx/zgQ69/ekem6IBDr/eESD+UgEOwfcJICB1AQ7DguEefNYBDsWkuSS
4YsEOyCIxHmZKAQ7J5BkiO52BDsvPCSLmLoEOzYPpJOAOAQ7PvLEgvscBDtGrOSHkegEOO2EJI8t5gQ7VGukhfhYBDtdGMSCm2QEO2YUpIRhtAQ7b0aEeYWYBDt3
bOSDWfgEO33bxH1oOAQ7hXlEh/eYBDuNoIR7bNAEO5cYpJhyXQQ7oMokkOyzBDuoAQSP4GoEO7ELpInSCAQ7tOMkj66CBDu+ziR3myAEO8XBhIvcggQ7OHxEhPxQ
BDvXduSHtzwEO93lBIlwbAQ76jREf+eABDvyf+SABhQEO/ipZIQDtAQ8ANSEh4dOBDwItyR/cDAEPBBFxH7MQAQ8F9ZEg4L8BDwghWSQ4PQEPCiwxH/leAQ8MKTE
jI1EBDw4XQStUyvEPEHkpKAYGcQ8SgakiTc+BDxRIcSAVqwEPFle5GyrYAQ8YCrkf/Y4BDxozgSItEQEPHCVRIZijAQ8eVQEht3gBDyAKwR6mMAEPIXV5Ig/ygQ8
jXIkjOwUBDyWSaSAMnwEPJ76xIEH5AQ8pTykgyCIBDytvsSDKBgEPLQpJH+riAQ8uRDEe4agBDzAFSSGiqQEPMdABIhR3AQ81NvEe844BDzeuQSAXKQEPOYNxI3e
SAQ88GAEenDQBDz5y4SGdIgEPQDLpIgQaAQ9DdgEfHPABDOXVkSDDIQEPR7fJGNEgAQ9JPEkf2vwBDOr8CSFWLwEPTLa5HEToAQ9Os7EeVZ4BD1IzASI4UQEPVE8
5IHwXAQ9WVOEfdrABD1iuOSLclAEPWohBI9lbgQ9cS9kfLhIBD16qySBYUgEPYNSpHOAIAQ9iijkfeioBD2RPMSAzRgEPZfsBIJbpAQ9oCOEgaSwBD2mGMSIDe4E
PaxCxHhd6AQ9sg+Ee17wBD28C4SD2WgEPcUNJIVFtAQ9ytokbJFgBD3YT+SLglQEPeNIhI49vgQ96DfEmW94BD3wEoSbrzeEPfhS5I/tOAQ+ADHki/uaBD4JGOSI
xAoEPg7f5HWO4AQ+FZDkeLvoBD4brySGBuQEPiW25IEkMAQ+LsgEcL+wBD437gSECcQEPkFn5HjbmAQ+SKqkCAAABD5SUASH6NQEPlfaJIsh+gQ+X6QkfYMwBD5m
mkSDB9gEPnAJBJQ3FAQ+d9rkgscwBD5+5iSLXgoEPoZMpHtBWAQ+j3Wkg7WkBD6WbaR7etAEPp3g5ITQiAQ+pjrEhWlUBD6teCRoKgAEPrQuBHkUkAQ+vGmkd42A
BD7CzCR7kaAEPsw3RIudHAQ+O/5keTCIBD7eEwSHzggEPunNZIUXGAQ+9pEkiwOmBD7/J6R6WWAEPwVaJI1spAQ/DXmEgl3UBD8WPGSG25AEPx5RBJBXyQQ/JKOk
gE1gBD8tnaSBnjQEPzcpJHmWQAQ/P3XkhWKkBD9M5ISD3hQEP1WwxHoykAQ/XOKkxIEgBD0mgOD4IwgED27C5HdSkAQ/eaVEgDpcBD+Dq+R78QgEP4j85HyD2AQ/
jymEfKBoBD+WzaR/S2AEP56GRHi9+AQ/qL3EgfEUBD+zBsR74RAEP7usZIj
```


A couple of screens of this

```
000.mzXML (0%)
```

Connected to mendel.duhs.duke.edu | SSH2 - 3des-cbc - hmac-md5 - none | 124x39 | 15, 39 | 00:09:33

2:mendel.duhs.duke.edu - [default] - F-Secure SSH Client

File Edit View Window Help

Quick Connect Profiles


```
Sjn7xG3wwARKPpmkeGrIBEpHWwRoGuAESksjhGVMgARKTwzEZhFABEpR2eRrYuAES1QCxI+ICARKVgXEhxSIBEpXhURxChAES1utZHXZYARKYFzkg+kUBEpiiGRs
VkAESmVtZGN9QARKZ99EfuhYBEppMyQEAAAESmzWZG6bYARKdVikgIUOBEp382R8HigESnmOJAQAAARKfayEiBKYBEqA8ARn2MAESoJMBH7HEARKg9BEex2YBEqF
YkR4KgAESonkpG1pYARKi91EeXWwBEqQOiR15IAESpODhIH9bARK1bZkar7gBEqYDeRqIAAESpqD5HQHIARKnaNkewKoBEqfmIR69EAESqPdhHMmYARKpRED+AAA
BEqnz+SKOkAESqqqZGj2oARKrSvkgMpQBEqxfWR2LeAESrW5pGn+IARKtx3EgbDoBEq6GWRw7aAESr3QRH5rOARKvztkiIOQBErA1SRvVkAESs1mBHXTgARKOGjk
jmsuBErRsuR1+AAEStLthIgglARK1IiD+AAABErbBIROk+AESt3ExFqtgARK4x8EcJdgBErklCR25/AESuXOBIiOygRK5OjkhLswBErrA4SDbggESuytRGNLgARK
7nskcxCQBErvsEQEAAAESvFgBGiEIARK+j7keEZoBEr8lyRvumAESv5eRGwVwARLAbXkdZygBEsDs4RrMkAESwUd5HKxQARLBpXkdAfgBEsIgWRsBUAESwoABAoA
AARLDEVkej2YBEsN6gRm2gAESw/aZHndIARLEjYEetmwBEsTsERjsOAESxxRZIgVUARLHkxEgKxcBEsgAWRuY2AESyJRxGzxoARLJ77keDDQBEspgKRbToAESyyU
5GMUgARLMMwkeJ9wBEszz2RxlQAESztjRI1a/ARLPJmEIYAABEtBHiRwyKAESOllRGtzIARLUqgEbwtABEtWJ+R7kwgES1/k5GQwgARLZaREZNjABEtrwIR5UrgE
S3BQJHRUIARLc5QEbMQgBEt5dQR3gsAES3vlZG16gARLf7qEdGvwBEuFsqR5F6AES4uxxH3NaARLjWZkaUngBEuRLcRfjYAES5SOhHIasARLmErkZq8ABEueEeRv
jkAES6SPZGkgYARLqmikWCKABEus5gRwnsAES7GTxHjDkARLstgkFwAABEu2MsR9lbAA==</peaks>
  </scan>
```

```
  <scan num="2"
        msLevel="2"
        peaksCount="56"
        polarity="+"
        scanType="full"
        centroided="1"
        retentionTime="PT1.900000S"
        startMz="95.0000"
        endMz="795.0000"
        lowMz="111.9760"
        highMz="781.0853"
        basePeakMz="149.0495"
        basePeakIntensity="2852380.0000"
        totIonCurrent="9755125.0000">
    <precursorMz precursorIntensity="6986078.000000"
                 collisionEnergy="35.000000">391.045410</precursorMz>
```


```
    <peaks precision="32">Qt/zvEchygBC4Xd+R6xyAELjErxGzCIAQxUMrkouGHBDFfXUR9cMAEMWlBJFj/gAQxsG4kaBGABDIHvgRoYEAEMh5sBFzRAAQy
Ri+EeVaQBDJtfAScsbOEMnuDBHjKUAQyixeEde+QBDRSG8RffoAENZ+cpGncIAQ1rMqkZv2ABDXjcwRqYMAENfBwRG1s4AQ2DoYEZ71ABDYeVCRsPKAEN9lVxF6D
gAQ39fXkbJ7ABDgJNcRanYAEOBbZxGzlAAQ4IVSEh/HOBDgmfySRwp4EOC25xHoZIAQ4Mu7kCAAABDiAasRyllAEOK4vxIHD6AQ4ttrkoCqKRDi/fASMIjwEOMWZ
hAgAAAQ5DSJEcAOgBDkTFkRvi4AEOR5LJGd6gAQ5RQPEdTxABDlVBgRqxGAEOWBJRIGIXAQ5a+qEYnIABDmb4kRxNQAEOjZsxGjOwAQ6/7kkXX6ABDs9LyRiGsAE
OOksxGpJYAQ7ZqZEawdgBDuQoeRha4AEO5OixHYN4AQ7pJ3kceYQBDuvvkR5XAgEO7VgRHidwAQ8JUjkaxtABDyrjWRrqqAEPLPZ5G1twAQ8uO1kaUHgBEQOV2Rh
3gAA==</peaks>
  </scan>

  <scan num="3"
        msLevel="2"
```


000.mzXML (0%)

Connected to mendel.duhs.duke.edu | SSH2 - 3des-cbc - hmac-md5 - none | 124x39 | 15, 39 | 00:09:59

2:mendel.duhs.duke.edu - [default] - F-Secure SSH Client

File Edit View Window Help

Quick Connect Profiles

```
  </scan>

  <scan num="2"
        msLevel="2"
        peaksCount="56"
        polarity="+"
        scanType="full"
        centroided="1"
        retentionTime="PT1.900000S"
        startMz="95.0000"
        endMz="795.0000"
        lowMz="111.9760"
        highMz="781.0853"
        basePeakMz="149.0495"
        basePeakIntensity="2852380.0000"
        totIonCurrent="9755125.0000">
    <precursorMz precursorIntensity="6986078.000000"
                 collisionEnergy="35.000000">391.045410</precursorMz>
    <peaks precision="32">Qt/zvEchygBC4Xd+R6xyAELjErxGzCIAQxUMrkouGHBDFfXUR9cMAEMW1BJFj/gAQxsG4kaBGABDIHvgRoYEAEMh5sBFzRAAQy
Ri+EeVaQBDJtfAScsb0EMnuDBHjKUAQyixeEde+QBDRSG8RffoAENZ+cpGncIAQ1rMqkZv2ABDXjcwRqYMAENfBwRG1s4AQ2DoYEZ71ABDYeVCRsPKAEN91VxF6D
gAQ39fXkbJ7ABDgJNcRanYAEOBbZxGzlAAQ4IVSEh/HOBDgmfySRwp4EOC25xHoZIAQ4Mu7kCAAABDiAasRyllAEOK4vxIHD6AQ4ttrkoCqKRDi/fASMIjwEOMWZ
hAgAAAQ5DSJEcA0gBDkTFkRvi4AEOR5LJGd6gAQ5RQPEdTxABD1VBgRqxGAEOWBJRIGIXAQ5a+qEYnIABDmb4kRxNQAEOjZsxGjOwAQ6/7kkXX6ABDs9LyRiGsAE
O0ksxGpJYAQ7ZqZEawdgBDuQoeRha4AEO50ixHYN4AQ7pJ3kceYQBDuvvkR5XAgEO7VgRHidwAQ8JUjkaxtABDyrjWRrqqAEPLPZ5G1twAQ8uO1kaUHgBEQOV2Rh
3gAA==</peaks>
  </scan>

  <scan num="3"
        msLevel="2"
mendel 223% tail 000.mzXML
    <offset id="3743">14882093</offset>
    <offset id="3744">14883275</offset>
    <offset id="3745">14885123</offset>
    <offset id="3746">14894120</offset>
    <offset id="3747">14896066</offset>
    <offset id="3748">14897268</offset>
  </index>
  <indexOffset>14898849</indexOffset>
  <sha1>62bbf6e59a640c22ec5d434708a11be4cc9758ba</sha1>
</msRun>mendel 224%
```

Finally we are at the end!

Connected to mendel.duhs.duke.edu | SSH2 - 3des-cbc - hmac-md5 - none | 124x39 | 21, 39 | 00:10:44

Discussion

- Data size: binary encoding
  - Good for data transfer
  - Bad for usage (querying, etc.)
  - Only small modification needed?
  - OPD: 021112.EcoliSol37.1
    - 87 MBs zipped
    - 136 MB unzipped

3 bytes becomes 4 chars

Wow these are similar!

Discussion continued

- Meta-data quality
  - Need further analysis
  - Comparison to MIAME, MIAPE, others?
- Documentation
  - Superb!

Tools

- Converters
  - Xcalibur native acquisition files to mzXML
  - MassLynx native acquisition files to mzXML
  - SCIEX/ABI Q-STAR instrument into mzXML
  - mzXML to sequest dta, mascot generic and micromass pkl formats
- Parsers
  - RAP, JRAP, RAMP (random access parsers)

Tools continued

- Viewers
  - mzXML viewer (Java)
  - InsilicosViewer (free, commercial-grade)

Databases

- Open Proteomics Database
  - http://bioinformatics.icmb.utexas.edu/OPD/#Links
  - Just a repository
- PeptideAtlas
  - http://www.peptideatlas.org/
  - SBEAMS
  - Just a repository
- The Global Proteome Machine (GPM)
  - http://gpmdb.thegpm.org/
  - Some searching capabilities – but I don’t understand

Datasets

- PNNL-NCRR
  - http://ncrr.pnl.gov/data/
  - Proteomics National Center for Cancer Research Resources
  - Capillary LC-MS/MS data
  - Proteomic investigation of *Deinococcus radiodurans* MR-1, gram-positive bacteria

Datasets continued

- ISB mzXML Data Repository
  - http://sashimi.sourceforge.net/repository.html
  - SASHIMI

Conclusion

- This is great
  - More tangible than MIAPI
- This needs work
  - Binary representation is not sufficient (or is it?)
- No effort to index?
- No open source databases?